

Claritas P\$YCLE® Premier 2021 Methodology

CONTENTS

- Introduction
- Data Sources
- Model Development
- Technical Support
- Legal Notifications

INTRODUCTION

Claritas P\$YCLE® Premier represents the next generation of P\$YCLE®. This segmentation system, created specifically for financial marketers, classifies every U.S. household into one of 60 consumer segments based in part on the income producing assets (IPA) of that household. P\$YCLE Premier offers an extensive set of ancillary databases and links to third-party data. This allows marketers to use data outside of their own customer files to pinpoint products and services that their best customers are most likely to use, as well as locate their best customers on the ground. This powerful segmentation system enables marketers to create a complete portrait of their customers by answering these important questions:

- Who are my ideal financial customers?
- What are they like?
- Where can I find them?
- How can I reach them?

P\$YCLE Premier’s external links allow for company-wide integration of a single customer concept. Beyond coding customer records for consumer targeting applications, Claritas provides estimates of markets, trade areas, and profile databases for behaviors. These behaviors range from leisure time preferences to shopping and eating to favorite magazines and TV shows, all of which can help craft ad messaging and media strategy. Components of the P\$YCLE Premier system can be grouped by the stage of customer analysis, as shown below:

CUSTOMER ANALYSIS STAGE	P\$YCLE® PREMIER COMPONENT USED
Coding customer records	Household-level coding
	Geodemographic coding and/or fill-in
Comparing coded customer records to trade area	Current-year segment distributions
	Five-year segment distributions

CUSTOMER ANALYSIS STAGE	P\$YCLE® PREMIER COMPONENT USED
	P\$YCLE® Premier Z6 segment distributions
Determining segment characteristics for demographics, lifestyle, media, and other behaviors	Claritas Neighborhood Demographic Profiles
	Claritas Household Demographic Profiles
	Claritas Financial Product Profiles
	Claritas Insurance Product Profiles
	Claritas Technology Behavior Track Profiles
	Claritas Income Producing Assets Profiles
	Claritas Net Worth Profiles
	Claritas Consumer Profiles
	Claritas TV Behavior Profiles
	Claritas Online Behavior Profiles
	Custom surveys or databases

DATA SOURCES

In developing Claritas P\$YCLE Premier, Claritas assembled a diverse set of financial and media behaviors from nationally representative data sources such as the proprietary Claritas Financial Track survey and Nielsen Scarborough. Each of these records included demographics and thousands of behavioral measures. The behavioral data included measures of both penetration and volume. For example, data is available for both whether a household owns a mutual fund (penetration) and about the current balance in that mutual fund (volume).

The P\$YCLE Premier model includes a new source of big data that offers property-level home value and property characteristics. The inclusion of this home value data in P\$YCLE Premier has resulted in a more predictive model that uses property value as a predictor of wealth and an array of financial behaviors. This data source provides a more precise read into wealthier segments, which are harder to identify and predict because they are a relatively small part of the U.S. population.

When implementing P\$YCLE Premier on third-party files, segment assignments depend on the third party's compiled list data. The unique models built for each third-party are designed to produce a distribution of assignments that mirror the distribution produced by the Claritas Multi-Source Aggregation and Distributional Alignment (MADA) process. MADA is a proprietary methodology informed by data from sources including Epsilon™ Data Management, LLC, Valassis™ Direct Mail, Inc., Infogroup™, and TomTom® North America, Inc. Such data includes, but is not limited to: age, income, and presence of children. This information is acquired from third-party providers who have a legal right to provide Claritas such information, and the data is either self-reported or modeled. This combination of data sources provides Claritas a unique competitive advantage in its segmentation assignment methodology, due to the unparalleled breadth and depth of address-level information. By combining data from multiple vendors with the Claritas Demographic Update, Claritas can make P\$YCLE Premier single assignments at the ZIP+6, ZIP+4, and ZIP Code levels, allowing better fill-in for records that do not get a household-level assignment.

The abundance of information available across multiple geographic levels allowed Claritas to construct a massive logical record for nearly 500,000 households. To each household record's name, address and behavioral data, Claritas added geographic identifiers at the census block group and ZIP+4 levels; evaluation characteristics that could be used to test and refine the segments covering the entire content of the survey and purchase data sets (providing thousands of profiles including both volume and penetration); and assignment characteristics that could be used to define the segments.

Examples include:

- Household-level demographics appended from the Epsilon TotalSource Plus™ file
- Neighborhood-level characteristics from the census block group information
- Summarized ZIP+4 and ZIP+6 level characteristics
- Claritas custom measures

The resulting database was then used to design and evaluate systems at four levels: household-level using self-reported data; household-level using list-based data; ZIP+4; ZIP+6 and census block group.

Note: All information was used for research only. Name and address fields were not retained for any third-party data after household-level demographics were appended to each record.

MODEL DEVELOPMENT

The methodology that now serves as the basis for Claritas P\$YCLE® Premier culminated years of research and development in a groundbreaking methodology that allows marketers to seamlessly shift from ZIP Code level to block group level to ZIP+4 level, all the way down to the ZIP+6 and individual household level—all with the same set of 60 segments. This single set of segments affords marketers the benefits of household level detail in applications such as direct mail, while at the same time maintaining the broad market linkages, usability, and cost-effectiveness of geodemographics for applications such as market sizing.

For decades, Claritas has set the standard for global market and consumer insight research. Our customer insights are based on representative samples of the population and help businesses understand what consumers watch, what they buy, and their financial preferences and behaviors to make your marketing more effective.

Claritas' segmentation solutions use a broad spectrum of demographic and lifestyle information to describe households and geography, enabling companies to better understand and anticipate customer buying behaviors. Our segmentation systems place each U.S. household into segments based on general consumer behavior and demographic characteristics. The segments are based on aggregated or modeled information that represent millions of households. No information about a unique individual or household is published or reported within segment assignments, making this a privacy safe solution.

New Statistical Techniques

Since the 1970s, the most popular of the clustering techniques has been K-means clustering. The final number of clusters desired is specified to the algorithm (this is the origin of the "K" in K-means) and the algorithm then partitions the observations into K-number of clusters as determined by their location in n-

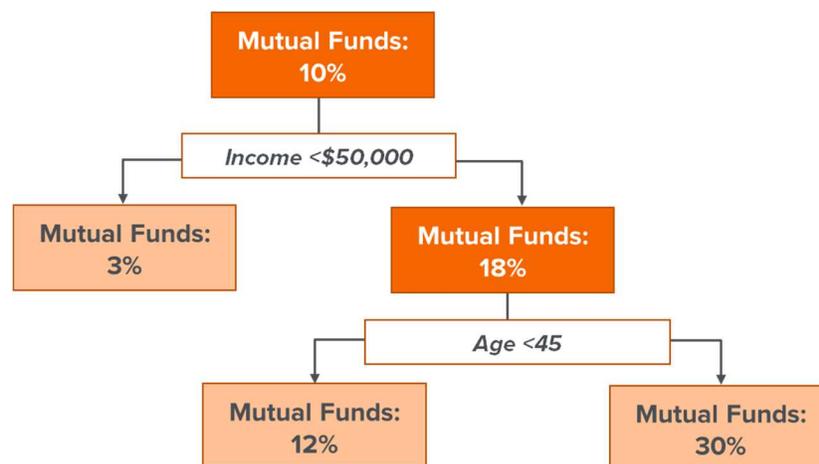
dimensional space, as dictated by demographic factors. Membership in a cluster is determined by the proximity to the group center, or mean, in space (hence the origin of the “mean” in K-means).

For any type of clustering process to work well, the statistician must correctly identify the important dimensions before implementing the clustering process. For marketing purposes, obvious drivers are age and income. However, appropriate levels for each of these critically important dimensions still need to be chosen. For example, does the dimension of income create better differentiation at \$35,000 or \$50,000? How does choosing between these two values of the same dimension change the clustering outcome? These choices are important, because when the clustering iterations end and yield an answer, marketers are left with clusters of households that have been organized by their proximity to each other by the demographic metrics that were chosen. This answer may or may not be meaningful to the original task of creating groups that differ in their behaviors—in large part because behavior measures were not incorporated into the clustering technique itself.

With P\$YCLE® Premier, Claritas broke with traditional clustering algorithms to embrace technology that yields better segmentation results. P\$YCLE Premier was created using a proprietary method developed by Claritas statisticians called Multivariate Divisive Partitioning (MDP). The MDP process borrows and extends a tree partitioning method that creates the segments based on demographics that matter most to households’ behaviors.

The most common tree partitioning technique, Classification and Regression Trees (CART), involves a more modeling oriented process than clustering. Described simply, statisticians begin with a single behavior they wish to predict and start with all participating households in a single segment. Predictor variables, such as income, age, or presence of children, are analyzed to find the variable—and the appropriate value of that variable—that divides the single segment into two that have the greatest difference for that behavior. Additional splitting takes place until all effective splits have been made or the size of the segment created falls below a target threshold.

In the example below, 10% of the entire household sample has a mutual fund. Using the demographic variable “Income < \$50,000” to create the first split of the original segment results in an optimized model with two segments—the first where 3% has a mutual fund and the second where 18% has one. If the 18% segment were split again, the optimized model would use the demographic variable “Age < 45” and result in two more segments—one with a mutual fund use rate of 12%, the other with a rate of 30%. This final model has three segments with usage rates of 3%, 12%, and 30%.



Note: The information above is purely to illustrate the technology used in segmentation and does not represent real data.

As this simple example points out, a significant limitation of the CART technique is that it generates an optimal model for only a single behavior. Because P\$YCLE® Premier is a multipurpose segmentation system, optimization across a broader range of behaviors is necessary. Claritas modifications to CART resulted in the MDP technique. This technique extends the simple CART process to simultaneously optimize across 250 distinct behaviors at once. This advancement allowed Claritas to take full advantage of the nearly 10,000 behaviors and hundreds of demographic predictor variables at different geographic levels, including household, that are available. The MDP process is run hundreds of times, varying the sets of behaviors, the predictor variable sets, and several other parameters to ensure that the resulting segments represent behaviorally important groupings.

New Assignment Data for P\$YCLE Premier

In addition to the geodemographic and behavioral data that was used in the development of the previous version of P\$YCLE®, there are two new, innovative features playing key roles in the new P\$YCLE Premier model. The first is property-level housing data for value and characteristics and the second is a technology score, which measures a household's use of technology in their daily activities. These two new measures influence the P\$YCLE Premier segment assignment for a household or geography.

The first new addition is a big data source of property-level home value and property characteristics from a third-party provider that covers more than 99% of U.S. properties. This dataset is sourced from tax assessor databases and includes various property-level attributes such as sales price, tax amount, total value, and the outstanding mortgage attached to the property (loan amount and duration). The inclusion of this home value data allowed Claritas to add a new high-end break to our Income Producing Asset Indicator dataset, which is a driver in P\$YCLE Premier. While Income Producing assets previously had 10 breaks and ended at \$2MM+, it now has 11 breaks and ends at \$3MM+. This allows for unprecedented distinction and definition of high wealth segments in P\$YCLE Premier.

The second new feature introduced with P\$YCLE Premier is a measure of technology use that identifies the extent to which a household has embraced technology in their everyday lives. A technology model was developed utilizing more than 100 technology related behaviors from several Claritas and third-party surveys. These behaviors include use of specific devices as well as specific activities engaged in by the household across various devices and channels. The technology use of each segment within the new P\$YCLE Premier system is described in terms of how the households within the segment scored relative to the average technology score. P\$YCLE Premier segments are described as High, Above Average, Average, Below Average or Low in terms of their use of technology.

Assessing the Role of Income Producing Assets

A distinguishing feature of P\$YCLE Premier is its use of the Claritas Income Producing Assets Indicators model, a proprietary Claritas model that estimates the liquid assets of a household based on responses to the Claritas Financial Track survey of financial behaviors—the largest financial survey in the industry—for which Claritas has actual dollar measures from each survey respondent, and permission to use this data for market research purposes. From the survey base, information for nearly 250,000 households (rolling three years of quarterly surveys) is anonymized, summarized, and used to construct balance

information for a variety of financial products and services that are core to income-producing assets (IPA). No individual respondent survey data is released with the P\$YCLE® Premier model.

Strongly correlated to age and income, IPA measures liquid wealth such as cash, checking accounts, savings products such as savings accounts, money market accounts and CDs, investment products such as stock and mutual funds, retirement accounts, and other asset classes that are relatively easy to redeem and move—and for which marketers can readily compete.

The Claritas Income Producing Assets Indicators model has been refined to use new data sources and data integration techniques. Claritas uses the 2016 Survey of Consumer Finances (SCF), a triennial cross-sectional survey of U.S. families conducted by the Federal Reserve Board, to adjust IPA distributions. Using the 2016 SCF, which has an oversample of wealthy families and a weighting scheme that corrects for under-coverage at top breaks, has enabled Claritas to provide accurate and stable estimates at the top of the IPA distributions. Coupling this with the addition of property-level home value, IPA breaks have been extended from \$2MM+ to \$3MM+, allowing for unprecedented distinction and a more focused understanding of high wealth segments.

These improvements to the IPA Indicators model allow the identification of seven IPA classes used to define and describe the P\$YCLE Premier segments.

- Millionaires
- Elite
- High
- Above average
- Moderate
- Below average
- Low

The values for these classes (except for Millionaires) will vary over time, to prevent arbitrary segment reassignments when markets rise and fall.

TECHNICAL SUPPORT

If you require further assistance, please contact the Environics Analytics support team between 9:00 a.m. and 8:00 p.m. (Monday through Friday, EST) at support@environicsanalytics.com or 888.339.3304.

LEGAL NOTIFICATIONS

P\$YCLE® is a registered trademark of Claritas, LLC. The DMA data are proprietary to The Nielsen Company (US), LLC (“Nielsen”), a Third-Party Licensor, and consist of the boundaries of Nielsen’s DMA regions within the United States of America. Other company names and product names are trademarks or registered trademarks of their respective companies and are hereby acknowledged.

This documentation contains proprietary information of Claritas. Publication, disclosure, copying, or distribution of this document or any of its contents is prohibited, unless consent has been obtained from Claritas.

Some of the data in this document is for illustrative purposes only and may not contain or reflect the actual data and/or information provided by Claritas to its clients.

Copyright © 2020 Claritas, LLC. All rights reserved.