CLARITAS 360

Custom Profile Basics

identify
deliver
optimize

Claritas 360

Custom Profile Basics

CONTENTS

INTRODUCTION

This document provides an in-depth explanation of the concepts associated with creating custom profiles from your customer address files. The process of transforming a customer address file into a segmentation profile can appear to be a daunting challenge. The kind of profile you create and the base you choose can impact your perception of what segments are more or less likely to use your product and what analysis areas provide the most potential.

WHAT IS A PROFILE?

A profile is a set of two frequencies. The first set is a series of counts (one for each segment) that represents the universe from which the behavior is drawn. The second frequency is a series of counts for those who "participated" in the behavior. Some profiles may include a third frequency representing consumption per segment, such as units consumed or dollars spent.

For example, if you profiled the "Bought Pretzels Last Week" behavior in a three-segment system, and you based the profile to households in your analysis area, the profile segment would look like this:

SAMPLE PROFILE		
SEGMENT	HOUSEHOLDS IN UNIVERSE (BASE)	HOUSEHOLDS WHO BOUGHT PRETZELS (BEHAVIOR)
1	1,523	25
2	3,246	72
3	957	13

Two Profile Types

A **standard profile** consists of only a count of behavior across segments and a base count across segments. Standard profiles are based on yes/no questions like "Did you drink cola last week?" The example above is a standard profile.

A **consumption profile** includes an additional measure—consumption by segment. This column can contain any measure of volume, including number of packages purchased, number of store visits, or account balance. It is important to remember that this number does not represent consumption by household, but aggregate consumption for the entire segment. In other words, total demand by segment.

For example, if "Bought Pretzels Last Week" was a consumption profile, it might look like this:

SAMPLE CONSUMPTION PROFILE			
SEGMENT	HOUSEHOLDS IN UNIVERSE (BASE)	HOUSEHOLDS WHO BOUGHT PRETZELS (BEHAVIOR)	NUMBER OF PACKAGES PURCHASED (TOTAL DEMAND)
1	1,523	25	78
2	3,246	72	234
3	957	13	52

For certain behaviors, consumption profiles give a very different picture than standard profiles. For example, if you asked "Did you drink cola last month?" you would probably find that nearly all segments answer "Yes" at a similar rate. However, if you ask "How much cola did you drink," you would find that certain segments consume far more cola than others.

Do You Need Profiles?

Profiles give you a picture of your product or service by consumer segment. With a profile, you can determine which segments are most likely to use your product or service and which segments make up the biggest part of your analysis area. This in turn allows you to determine which segments to pinpoint, which are prime for development, and which should be avoided in your marketing efforts.

The Source of Behavior Counts and Consumption Data

Behavioral counts come from a wide variety of sources, including but not limited to surveys, direct measurements of behavior, demographic data, geographic areas, and your own data.

Surveys

Companies like us, The Nielsen Company (US) LLC (Nielsen), Simmons Media Research Bureau (Simmons), and GfK Mediamark Research &Intelligence LLC (MRI) collect large amounts of survey data. Because these surveys contain the name and address of each individual who completes the survey (and often household-level demographics), we can assign a segmentation code to each response. These responses are then summarized by segment code to create behavioral counts.

Direct Measurements of Behavior

Entities such as Nielsen's media research arm do not survey their panel households. Instead, they measure their behavior directly – an electronic device records their television viewing habits. Again, because the address of each participating household is known, we can assign a segmentation code to each response and produce behavioral counts. Nielsen NetRatings follows a similar procedure. Their panel households install software that records the Internet sites they visit. This allows profiles of Internet usage to be built.

While we, and many of our trusted third-party data providers, collect data on a panel (or representative sample) of households, other companies measure the behavior of all or nearly all the households involved in a given activity. For example, R. L. Polk & Co. directly measures automobile ownership by collecting data from the various state motor vehicle administrations. In this case, actual household-level data, such as auto registrations or insurance policies, is never released. It is either summed by segment or summed geographically at the ZIP+4 level or higher.

Demographic Data and Geographic Areas

Behavioral counts can also be created using demographic data. The behavioral count may be the number of people who fall into a certain age group (by segment) or the number of people who own a home.

SAMPLE DEMOGRAPHIC PROFILE		
SEGMENT	HOUSEHOLDS	OWNER-OCCUPIED HOUSEHOLDS
1	1,523	25
2	3,246	72
3	957	13

Claritas demographic profiles are available at the household and neighborhood level. When the variable is total households, you are said to be profiling the geographic area.

SAMPLE GEOGRAPHIC PROFILE		
SEGMENT	U.S. HOUSEHOLDS	ANDERSON, SC CBSA
1	62,624,621	72,314
2	43,432,987	36,128
3	3,891,620	2,641

Your Data

The most valuable source for behavioral profiles is your own data. Appending segmentation codes to your file and summarizing by segment produces a distribution of your customers across segments. Summing by geography can produce actual penetrations and composition numbers by geographical unit, such as ZIP Code.

WHAT IS A BASE?

A base is the universe from which the behavioral count is drawn, or the "loaf" from which the behavioral count is "sliced." For example, if the behavior is "Subscribe to the Daily Times," the base may be total households. If the behavior is "Bought a new SUV," then the base may be total new car buyers.

Without a base count, you can only determine what percentage of those engaged in the behavior fall in each segment. This measure, commonly called percent composition or % Comp, tells you the relative size of the segment (quantity), but gives no indication of the likelihood that the segment will express the behavior in question, because you have no relative benchmark to compare the size of those doing the behavior to their presence in the analysis area.

MEASURES AVAILABLE WITHOUT A BASE		
SEGMENT	BOUGHT PRETZELS LAST WEEK	% COMPOSITION
1	25	22.73
2	72	65.45
3	13	11.82
Total	110	100

With a base count in place, you can determine what percentage of all possible households in a given segment express the behavior in question (percent penetration) and how **likely** it is that a segment will express that behavior verses the overall average (index). Penetration and Index are measures of **quality**.

MEASURES AVAILABLE WITH A BASE						
SEGMENT	% HHS IN UNIVERSE (BASE)	% COMP	HHS WHO BOUGHT PRETZELS LAST WEEK (BEHAVIOR)	% COMP	% PEN	INDEX
1	1,523	26.6	25	22.73	1.64	85
2	3,246	56.7	72	65.45	2.22	115
3	957	16.7	13	11.82	1.36	71
Total	5,726	100	110	100	1.92	100

In the table above, there are 1,523 households in the analysis area that belong to Segment 1. They make up 26.6% of total households in the analysis area.

(1,523 / 5,726) X 100 = 26.6

Twenty-five of those Segment 1 households, or 1.64%, bought pretzels last week. Of all the households who bought pretzels last week, 22.73% were from Segment 1.

(25 / 1,523) X 100 = 1.64

(25 / 110) X 100 = 22.73

Segment 1 households are 0.85 times as likely, or 15% less likely, to have bought pretzels last week than the average household.

(22.73 / 26.60) X 100 = 85 OR

(1.64 / 1.92) X 100 = 85

100 − 85 = 15

Choosing the Right Base

Choosing an illegitimate base is the most common mistake in the profile creation process. Incorrect bases lead to incorrect penetration and index scores, which ultimately leads to poor decisions, concerning which segments to pinpoint, and incorrect market demand estimates.

Four important rules will help you to avoid choosing an incorrect base:

- The base must be created using the same segmentation system and level as the behavioral count.

- The base must include all of the behavioral count that falls in the base analysis area.

- The base should be made up of those individuals or households that have the opportunity (if not the desire) to purchase the product.

- Choose the base with the purpose of your analysis in mind.

Step 1: Create the base using the same segmentation system and assignment level as the behavioral count

The count of households in a given area can vary greatly between data providers and segmentation products. These variations usually result from different definitions of "household" and the variety of source data inputs. For this reason, be sure to use the same data provider and segmentation level to tally base counts as you do to tally behavioral counts, as follows:

- If you are using a geographic-level system, use Claritas geographic-level base counts, often designated as "CY" (for current year).

- If you are using a ZIP+6 level system, use Claritas household level base counts because they are very closely aligned with ZIP+6 level assignments.

- If you are profiling your own data, you may base to either Claritas counts (for the level at which the data was assigned a segment), or to a base derived from the same database. For example, if you were profiling respondents to a direct mail campaign, you could base respondents to number of pieces mailed. If you were profiling a particular product, you might base the count of customers who bought that product to your total customers.

- If you are profiling survey results, base respondents who answered "yes" to a given question to total respondents.

Notice the variation in household counts for the levels listed below.

HOUSEHOLD COUNT FOR ZIP CODE 22201	
CLARITAS UPDATE (GEO-LEVEL) (CY)	CLARITAS HOUSEHOLD LEVEL (HH)
15,164	17,077

Compiled list data is different than Claritas Pop-Facts data because compiled list data is rooted in direct marketing lists while Pop-Facts® data is rooted in the decennial census. For example, a list may count two unmarried adults living at the same address as two households, while the census would count them as only one household. Other variations have to do with coverage and most often occur in younger, urban areas and in rural areas. Young urban areas show these variations due to factors like high mobility, a larger number of unmarried couples and a large number of housemate situations. Variations in rural areas can be attributed to less than complete coverage in the compiled list files that drive the household assignment.

This variation in household counts can result in great variation in index scores. The following index scores are for "Bought a New SUV," with data collected at the household level but appended with a ZIP+4 segment assignment.

INDEX VARIATION "BOUGHT A NEW SUV" PROFILE CREATED WITH CLARITAS PRIZM PREMIER Z4		
SEGMENT	CLARITAS UPDATE BASE (CORRECT)	CLARITAS HOUSEHOLD LEVEL BASE (INCORRECT)
04 Young Digerati	55	175
07 Money & Brains	138	135
26 Home Sweet Home	226	182

These variations are due to misalignment of the appended segments (at the ZIP+4 level) versus two sets of base counts—one driven by ZIP+4 segments (which is correct) and the other by household lists, which in this example is incorrect.

Step 2: Ensure that the resulting profile includes only the behavioral count that falls within the base analysis area

Determining a proper base requires a good understanding of how, where, and to whom the product is offered.

In the following diagrams, think of a set of customers as being represented by the smallest circle. The larger circles represent both proper and invalid base possibilities. For a base to be proper, the smaller circle (behavioral count) must be completely surrounded by the larger circle (base). Any time the smaller circle covers an area not covered by the larger circle, the base is invalid.

Note: Some MRI profiles provide counts of adults. In order to ensure accurate projections to household ground counts, these profiles are based to households. In this situation, percent penetration may be greater than 100, and should be interpreted as "users per 100 households."

PROPER BASES	INVALID BASES	
The **smallest** circle represents the customers of a mail order catalogue. The **middle** circle represents the households that receive the mail order catalogue. The **biggest** circle represents all households in the U.S.	The **smaller** circle represents the customers of a mail order catalogue mailed only to New England states. The **larger** circle represents all households in the state of California.	The **smaller** circle represents the customers of a mail order catalogue in Southern and Northeastern states. The **larger** circle represents households in Southern states, and does not include any households in Northeastern states.
-or- The **smallest** circle represents the customers of a bank branch. The **middle** circle represents all households within five miles of the bank branch. The **biggest** circle represents all households in the U.S. While the largest circle, which represents all U.S. households, is not wrong, it is also not the most appropriate base. The narrower definition (the middle circle) would yield a better result.	-or- The **smaller** circle represents all subscribers to a cable television system in Seattle. The **larger** circle represents all the households in Los Angeles.	-or- The **smaller** circle represents all males in the U.S. between 25 and 34 years old. The **larger** circle represents all males in the U.S. over age 30.

For some applications, it may be better to split the customer file into subsets and create unique bases and profiles. For example, an amusement park might create one profile of day-trip customers within a three-hour drive of the park and another profile for those traveling from farther away. Each set of customers would have a different base—the first for households within a 200-mile radius and the second for a larger region made up of several states or the entire country. It is important that only customers within 200 miles be included in the first profile (to match the base), while all visitors could be counted in the larger base.

Step 3: Ensure that the base is made up of those individuals or households that have the opportunity (if not the desire) to purchase the product

"Opportunity" is a broad word that requires further explanation. For our purposes, individuals or households who have the opportunity to purchase a product have geographic and legal access to the product.

Geographic Access

To focus your attention on the places where your product is the most accessible and the most successful, examine the compositions, penetrations, and indices for the different areas where your customers live. The general rule of thumb is that your base should include at least 80 to 90 percent of all customers.

A penetration report like the one below can help you determine your base area.

ZIP		ALL SEGMENT HOUSEHOLDS		CUSTOMER HOUSEHOLDS				
CODE	NAME	COUNT	% COMP	COUNT	% COMP	% CUM	% PEN	API
85254	Scottsdale	19,533	14.68	8,360	29.28	29.28	42.80	322
85251	Scottsdale	18,468	13.88	6,216	21.77	51.05	33.66	253
85263	Rio Verde	883	0.66	199	0.70	51.75	22.54	169
85260	Scottsdale	16,059	12.07	3,190	11.17	62.93	19.86	149
85258	Scottsdale	12,318	9.26	2,383	8.35	71.27	19.35	145
85253	Paradise Valley	7,364	5.53	1,365	4.78	76.05	18.54	139
85257	Scottsdale	13,607	10.22	2,222	7.78	83.84	16.33	123
85259	Scottsdale	11,471	8.62	1,787	6.26	90.10	15.58	117
85262	Scottsdale	4,191	3.15	526	1.84	91.94	12.55	94
85255	Scottsdale	10,013	7.52	1,200	4.20	96.14	11.98	90
85256	Scottsdale	1,743	1.31	193	0.68	96.82	11.07	83
85268	Fountain Hills	9,300	6.99	890	3.12	99.94	9.57	72
85264	Fort Mcdowell	293	0.22	17	0.06	100	5.80	44
85643	Willcox	3,192	2.40		0	100	0	0
85631	San Manuel	2,083	1.57		0	100	0	0
85623	Oracle	1,929	1.45		0	100	0	0
85618	Mammoth	609	0.46		0	100	0	0
85545	Roosevelt	31	0.02		0	100	0	0
Total		133,087	100	28,548	100	100	13.31	100

The highest penetration of customers is in ZIP Code 85254 at 42.80%. This ZIP Code accounts for 29.28% of all the customers. (With an API—actual penetration index— of 322, it is 3.22 times greater than the average penetration rate of 13.31.) Moving down the list, we find that ZIP Code 85251 has an above average penetration (index of 253) and contains 21.77% of all customers.

Using the general rule of 80%, the cutoff is between ZIP Codes 85253 and 85257 (a cumulative (%Cum) measure that provides a running tally of the percent composition rate, which in this case is between 76.05% and 83.84%).

Other Access Issues

Specific items to consider when determining which individuals or households have the opportunity to purchase your product are illustrated in the following table.

OPPORTUNITY CONSIDERATIONS	
Accessibility	Is the product more accessible, less accessible, or not accessible in the analysis area? Is accessibility a result of regional appropriateness of distribution trends (i.e., snow shovels not being sold in Miami)? Is the technology that delivers the products available in the area (i.e., cable TV or broadband Internet)?
Trade Area	If you have retail-based distribution, how far will customers travel to a store? Is there a difference in how far a customer will travel based on the population density of the store area, such as urban versus rural? Is the trade area national, regional, or local? Do regional or local trade areas create biases (i.e., rural segments not found in New York City)?
Distribution Channels	Is the product retail-delivered, so that the customers must come to a store, or can they order from a catalog? To whom do you send the catalog? Can customers order over the Internet?
Population Verses Household Bases	Do individuals use the product or is it consumed by several or all the people in the household? Population and adult counts can be based to household counts. In this case, percent penetration becomes "users per 100 households."
Legal Limits	Is there some legal limitation on the users of your product (i.e., automobiles, alcohol, and tobacco)?

Step 4: Choose the base with the purpose of your analysis in mind

Profiles are developed in order to compare the behavioral count to the base. For this reason, it is important to keep the purpose of your analysis in mind when choosing a base.

For example, you would not want to base "Bought a New SUV" to households with an income of $50,000 plus. If you did, you would eliminate a number of segments who are not likely to buy any new SUV, and therefore lower the index scores for segments who are likely to buy an SUV. Also, because you cannot guarantee that no one with an income of less than $50,000 bought an SUV, base creation Step 2 (above) is violated.

If your intention is to perform segment analysis or to compare profiles, basing "Bought a New SUV" to all households would yield accurate information, but would show less differentially than you might desire because the profile of all new-vehicle buyers is so upscale. In other words, you would see that only very high-income households buy new SUVs, but this is largely true because high income households are much more likely to buy new vehicles to begin with. The analysis would be technically correct, but it might not provide the information you are looking for. It may be more useful to base "Bought a New SUV" to "Total New Vehicle Buyers." This would show you the

difference in SUV buyers and the vehicle-buying public and might be most useful in competitive analysis.

However, if your intention is to create market projections, it is necessary to base "Bought a New SUV" to households. Market Potential calculations rely on households and, as such, can only be used with profiles based to households. This is also true for media and lifestyle rankings because these surveys are typically representative of all households, not just new vehicle buyers.

Types of Bases

Bases fall into five broad categories:

- National
- Regional
- Local
- Aggregated Trade Areas
- Specialized

National Bases

National and super-regional bases are appropriate for companies whose products are available and distributed nationwide or nearly nationwide, and whose products are generally available to all who want to purchase them. **National** profiles include behavior counts of all customers for a product or service, and base counts for the entire country. National bases are easy to update and very rarely absolutely "wrong," but at times other kinds of bases are more appropriate.

Super-regional profiles include behavioral counts and base counts for broad regions (for example, states east of the Mississippi), but not the entire nation.

Regional Bases

A common mistake is to assume that products that are available nationwide (to anyone with the desire and means to buy them) should be analyzed with a national base. Many times, the actual distribution pattern for these products shows that they have virtually no sales in a large portion of the country.

For example, a high-end auto manufacturer may start with a national base because their cars are available in every state. For many applications, this is the right choice. But a closer examination might indicate that more than 80% of auto sales come from the Southwest, the South and the East Coast. In this case, creating a superregional profile using behavioral and base counts from these regions may be more appropriate.
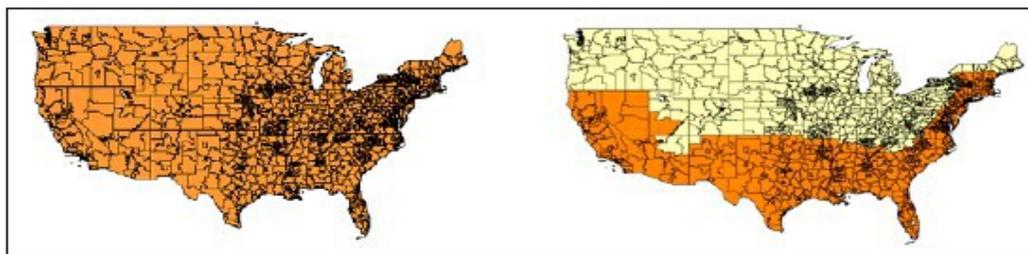
National profile (on left) and regional profile (on right)

Local Bases

Local profiles are appropriate for organizations whose distribution is centered around a single trade area or a small number of trade areas. Users create local profiles by determining the penetration of customers in small geographies (county, three-digit ZIP Code, or below). The group of geographies in which 80% to 90% of customers reside are then mapped. "Holes" are filled in and extensions are trimmed to create a contiguous trade area from which to draw both behavioral and base counts. In general, "tighter" base areas are better than "looser" base areas, even if less than 80% of customers are included.

Choosing a Level of Geography

The level of geography at which you build a profile is determined by the size of your trade area. In general, larger trade areas are built from higher levels of geography. Large trade areas may be expressed in counties or three-digit ZIP Codes. Smaller trade areas may be expressed in ZIP Codes, census tracts, or block groups.

Mapping Your Area

Mapping your trade area allows you to fill in "holes" and trim extensions where you might not have customers currently, but where customers have geographic access to your product. Base counts are then produced for this new area, which become the universe from which behavioral counts are drawn.
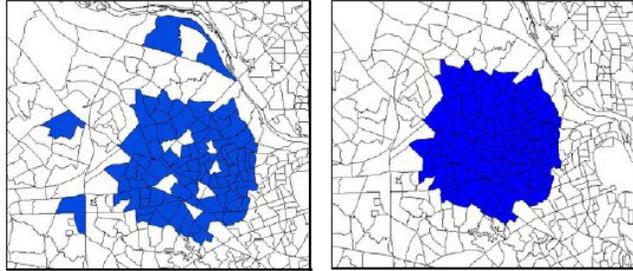
Trade area with holes and outliers (on left) and final trade area (on right)

Aggregated Trade Area Bases

An aggregated trade area base is simply the combination of the many local base areas that surround individual locations. They are appropriate in situations where multiple locations exist, but customers have to come to a physical location to make a purchase. The local base areas can be uniform (all 2-mile radii), variant (1/4-mile radii for city locations and 3-mile radii for suburban locations) or custom (individually defined trade areas for each location). Common trade area types are radii, polygons, or drive- time polygons. Trade areas are aggregates of the geounits that fall within these shapes.
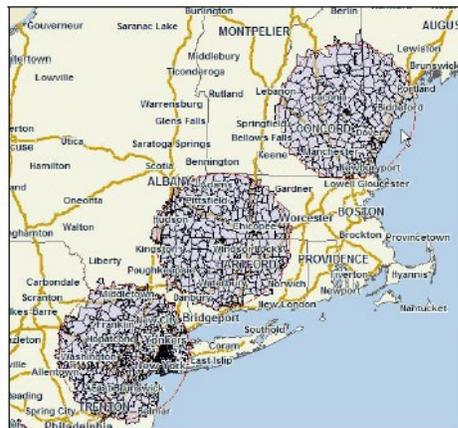
Aggregated trade areas

Specialty Bases

Specialty profiles consist of behavioral counts and base counts created from a company or survey database. In other words, the base count is not households. For example, you might want to base total buyers of a particular product to all customers, or total respondents to a mail campaign to pieces mailed.

To create a specialty profile, summarize both the behavioral count and the larger base count by segment.

Note: Market Potential calculations in Claritas software rely on household counts. Therefore, specialty profiles based to something other than households may produce inaccurate market potential indices.

Choosing Your Base Type

To determine if a national base is appropriate, start by summarizing your data by large area geography like State or DMA. If you have a large number of customers in nearly every large geography, or a super-regional collection of large geographies, a national base is most appropriate. If this is not the case, resummarize your data at the County or 3 Digit ZIP level to determine whether a pattern of geographic distribution becomes apparent. If so, a regional base may be appropriate. If a concentrated pattern of distribution is apparent at these levels, your base may be local or determined by trade areas and you may wish to summarize at a lower level within those concentrated areas.

How Base Selection Affects Indices and Penetration

National bases tend to yield higher indices, but lower percent penetrations because areas with high concentrations of segments with little or no penetration are included. Thus, high-buying segments look better by comparison, hence the higher index scores.

Local and specialized bases tend to yield flatter (but possibly more accurate) indices and higher percent penetrations. As the base narrows, areas with high-concentrations of segments with little or no penetration are excluded. The remaining area has a higher concentration of highly penetrated segments, driving the indices of the top segments down by comparison. Because areas with few or no customers are excluded, percent penetration increases.

BASE COMPARISON FOR THE "MOVERS AND SHAKERS" SEGMENT					
BASE	**BASE COUNT**	**BEHAVIOR COUNT**	**% PEN**	**TOTAL PROFILE % PEN**	**INDEX**
National	1,805,793	39,593	1.97	0.36	537
Regional	1,716,389	34,631	2.02	0.48	410
Trade Area	1,113,009	26,002	2.34	0.59	385
Local	58,037	1,483	2.55	0.76	336
Specialized	169,118	35,593	21.04	21.04	294

For example, mass transit riders index high for urban segments and even higher if non-mass-transit accessible areas in the exurban areas are included in the base.

For many profiles, the same pattern will emerge with different bases. In this case, the national profile generates sharper indices than the regional profile, but the general pattern is the same. Both profiles are useful. The national profiles can be used for national media strategy, but the

regional profile may be more ideal for estimating market demand for the analysis areas in the base region.
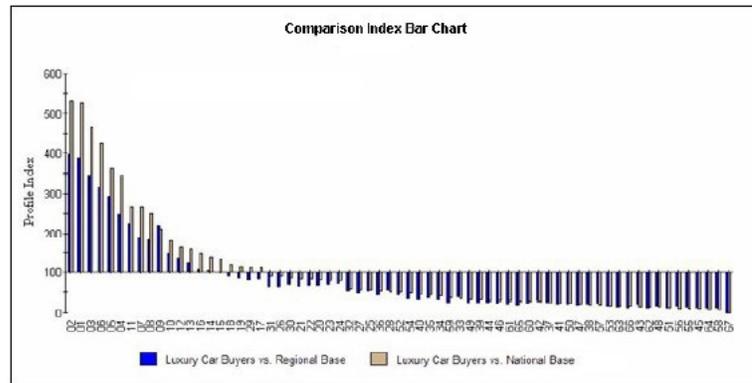
National- and regional-based profiles

Summary

A profile is a set of two frequencies comparing those who participated in a particular behavior to the universe from which that behavior is drawn. Standard profiles consist of a count of behavior across segments and a base (universe) count across segments. Consumption profiles include an additional measure—total consumption by segment.

Profiles give you a picture of your product or service by consumer segment and allow you to determine which segments to pinpoint, which are primed for development, and which should be avoided in your marketing efforts. Profiles are created from surveys, direct measurements of behavior, demographic data and geographic areas, and your own data.

Choosing a base is key to the profile creation process. Bases must:

- Be created from the same segmentation system and assignment level as the behavioral count
- Include only the behavioral count that falls in the base analysis area
- Be made up of those individuals who have the opportunity to purchase the product
- Be chosen with the purpose of the analysis in mind Bases fall into five broad categories:
- National bases are appropriate for companies whose products are available and distributed nationwide or nearly nationwide, and whose products are generally available to all who want to purchase them. While a good starting point, and almost never wrong, national bases also may not be the most accurate or appropriate.
- Regional bases are used for those companies whose actual sales patterns show concentration in a region of the country, for example, states east of the Mississippi.
- Local bases are appropriate for companies whose distribution is centered around a single trade area or a small number of trade areas.
- Aggregated Trade Area bases are combinations of many local base areas surrounding specific locations, and generally they yield the most accurate percent penetration rates.

- Specialty bases consist of behavioral counts and base counts created from a company database. In other words, the base count is not households.

The base you choose affects indices and percent penetration. The larger the base is, the higher the index scores, but the lower the percent penetration. As the base narrows, index scores flatten, and percent.

LEGAL NOTIFICATIONS

Pop-Facts and PRIZM are registered trademarks of Claritas, LLC. 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.