

Segmenting with the PPI®

A GUIDE TO USING THE PPI IN MARKET SEGMENTATION

Pro-poor organizations use client segmentation for market research, customer profiling, and evaluating the effectiveness of their products or services. The process of segmentation involves dividing an organization's target market or existing clients into groups that have common needs, behaviors or other characteristics. Doing so allows the organization to design and implement segment-specific strategies that more appropriately and effectively serve these groups.

Segmenting using poverty data

Ensuring that poor households have access to the appropriate knowledge, services, and products is fundamental to pro-poor organizations' missions. Effectively doing so requires research and evaluation, of which segmentation can play a significant role.

Better understanding how poverty affects the way that your beneficiaries interact with products, services and interventions allows management ultimately to make better decisions around making improvements for the clients' benefit.

Market research

Market research permits a deeper understanding of the environment in which an organization is working. The goal of market research is to identify segments of individuals with one or more characteristics that cause them to have (or are correlated with) relatively similar product needs. These needs can then be researched further through surveys and client-level information.

Let's use a hypothetical example to show how segmenting using the PPI can improve targeting and strategy. Suppose an organization plans to provide a payment service via a mobile app to individuals living under the \$1.25/day poverty line. A fundamental requirement for access to this service is access to a mobile phone. By using the PPI to segment the population prior to the program's launch, along with data on the possession of a cell phone for each household, the organization received the results shown in Table 1, indicating that only 12.5% of their target market have access to a cell phone. (See *Segmenting by Poverty Lines* below for a description of the method used to calculate this information.) Collecting and analyzing the data prior to the launch of the program will save the organization considerable time, effort, and funds.

Table 1: Market Research*

Poverty Level	% of Clients with Mobile Phone
Under \$1.25/day poverty line	12.5%
Between \$1.25/day and \$2.50/day poverty line	20.3%
Above \$2.50/day poverty line	67.2%

Evaluating products and services

Product and service evaluation aids organizations in better understanding the impact of their programs, the quality of their services, and the degree to which they are reaching the right clients. Segmentation deepens this analysis because it can indicate how different types of clients respond differently to an organization's product offerings.

Using this information, management can decide whether adjustments should be made to increase usage among the target market with an aim to deliver the right products and services to the right clients.

A microfinance institution (MFI) in the Philippines used poverty segmentation to evaluate the uptake of a new savings product meant for the poorest. Management was surprised to see that poorer customers were less inclined to use the new product. Through follow-up focus group discussions, they realized that these poorer customers were not using the product because the minimum opening balance was too high. After gaining this insight, the MFI lowered the minimum opening balance and saw not only that account openings significantly increased, but poverty outreach—the percentage of clients living below the national poverty line—also increased by 7% within the first few months. In addition to using PPI data to understand the adoption of the product among varying poverty levels, the MFI was able to validate that the product design change did have an effect on uptake among the poorest.

Three Segmenting Methods

There are three ways to use the PPI in market segmentation. The first method groups individuals using data other than the PPI—such as gender, geography, and products used—and then uses the PPI to understand differences in poverty levels among these groups. The second and third methods define segments using PPI data alone. Of these two, the first uses poverty likelihoods or PPI scores to segment and the second uses poverty lines.

This guide introduces segmenting methods in order of increasing complexity. If you are just beginning to segment your clients and customers, the first and second methods are sufficient to help management uncover insights that lead to change. The third, segmenting by poverty lines, is complex, but the advantage is that differences in client behavior and characteristics may be linked to poverty lines, e.g., “Clients living between the \$1.25/day and \$2.50/day poverty lines use our new service more than any other segment.”

For all of these methods, it’s important to keep in mind that segments should be different enough from one another that their distinction makes sense and also large enough to be both statistically significant and meaningful for business decisions. However, segments needn’t be the same size.

Segmenting by non-PPI Population Characteristics

We’ll start by discussing how to use characteristics of the population other than PPI data as the distinction used to create segments. Segmenting clients in this way permits a clear analysis of the poverty characteristics of each group. This method is currently the most commonly employed by PPI users and is also the easiest. First, using some variable, such as gender, a population is divided into groups, such as male and female. The poverty outreach, given in terms of one or more poverty lines, of these groups is then calculated.

The hypothetical example below shows how a client population was split by the following occupations: merchant, crop farmer, and livestock farmer. Then the distribution of the occupations by poverty level was calculated. This information would help a pro-poor organization that wanted to serve those most likely to be poor to focus their efforts on livestock farmers.

Table 2: Segmenting by Occupation

Occupation	Under \$1.25/day Poverty Line	Between \$1.25/day \$2.50/day Poverty Lines	Above \$2.50/day Poverty Line
Merchant	11.2%	32.3%	56.5%
Crop farmer	45.5%	25.4%	29.1%
Livestock farmer	52.3%	33.5%	14.2%

Table 2 is an example of how to begin thinking about segmenting by other characteristics, but there are many more possibilities for analysis. Segmenting by products or services can help an organization understand which products are reaching the poor. The following example shows how an organization can cluster clients by product (assuming clients can only have one product):

Table 3: Segmenting by Products and Services

Product	Total Number of Clients	% Below National Poverty Line	Total Number of Poor Clients
Product A	1,563	16.8%	262
Product B	37,276	21.5%	8,014
Product C	85,555	10.3%	8,812
Product D	4,203	12.4%	521

This table shows that the poorest people (defined as those living below the \$1.25/day poverty line) are accessing or using product C, even though product B has a higher poverty concentration. An organization may then ask itself, “Why do so many poor clients choose product C?” This could then inform strategy during the organization’s next phase of product implementation.

The most common characteristics used to segment with this method are gender, location and occupation; however, as the example above shows, there are many other variables that can be used in this type of analysis. Anything that can be used to group a population, from demographic to financial to transactional data, may be used for segmenting. It’s up to data analysts, along with management, to decide which variables are most likely to lead to interesting insights about their clients.

Segmenting by Poverty Likelihood or PPI Score

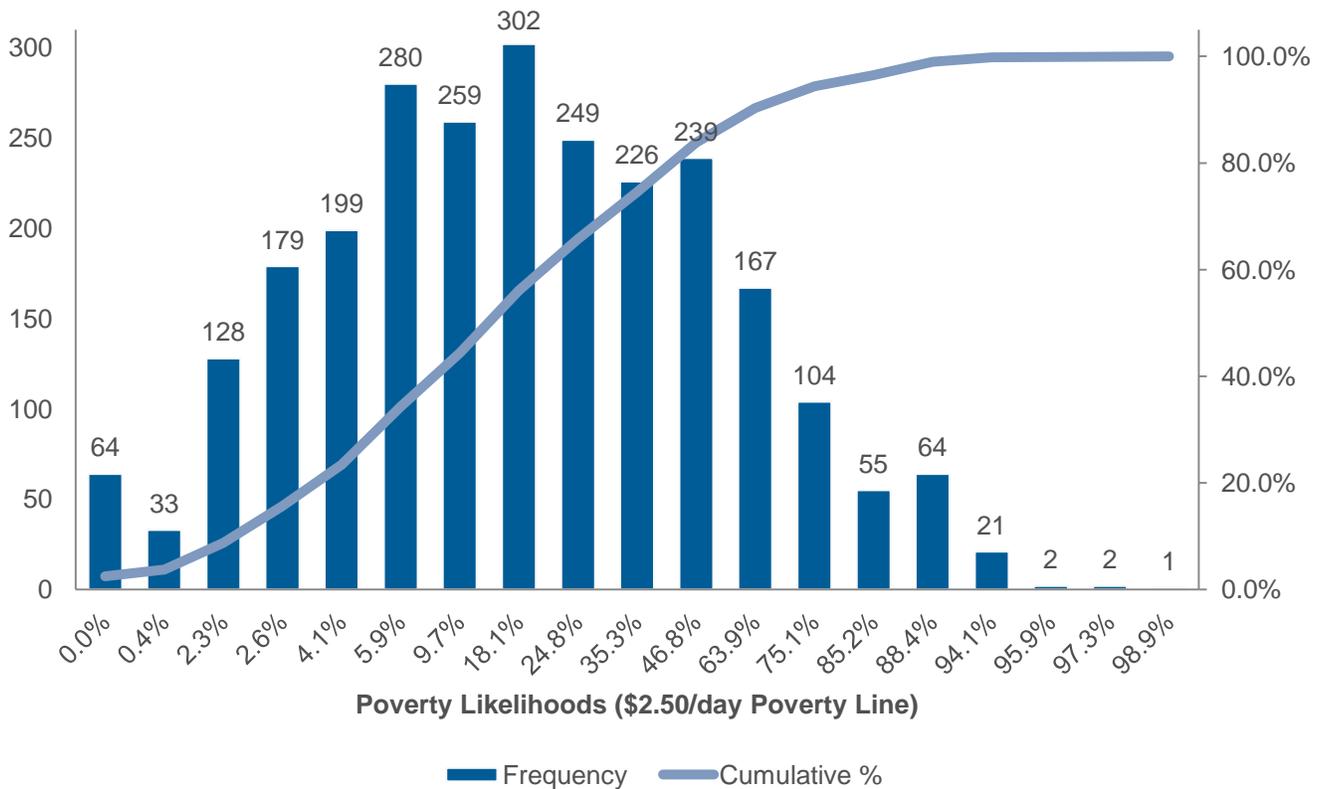
The above method used client characteristics to segment a population and then assessed the poverty levels of each segment. The method in this section also uses client characteristics, but the order of analysis is switched. First, the PPI is used to segment a population, and then non-PPI characteristics of each segment are analyzed.

The most basic way to segment a population using the PPI is by PPI score or PPI likelihood. Because each PPI score is associated with one (and only one) PPI likelihood, the two methods are equivalent except for the units in which the segments’ cut-offs are defined. The strategy is to assign clients to segments based on their PPI scores or PPI likelihoods and then use other information on these clients to characterize those segments.

Example

Refer to the data in Chart 1 from a social enterprise in the Philippines representing 2,574 beneficiaries, which shows the distribution of poverty likelihoods, as well as the cumulative percentage of clients for each likelihood:

Chart 1: Segmenting by Poverty Likelihood



Choosing the ranges of likelihoods that define a given segment is up to the analyst, but remember that each segment should be large enough that differences across segments are likely to be real and not due to a bad luck-of-the-draw using a small sample, as well as different enough from the others to be meaningful for analysis.

Table 4: Cumulative Percentages for PPI Likelihoods

Poverty Likelihood	Frequency	Cumulative %	Poverty Likelihood	Frequency	Cumulative %
0.0%	64	2.5%	46.8%	239	83.8%
0.4%	33	7.5%	63.9%	167	90.3%
2.3%	128	8.7%	75.1%	104	94.4%
2.6%	179	15.7%	85.2%	55	96.5%
4.1%	199	23.4%	88.4%	64	99.0%
5.9%	280	34.3%	94.1%	21	99.8%
9.7%	259	44.4%	95.9%	2	99.9%
18.1%	302	56.1%	97.3%	2	100.0%
24.8%	249	65.8%	98.9%	1	100.0%
35.3%	226	74.6%			

For our example, we will divide the data roughly into thirds and keep each segment comparable in size at around 860 clients per segment (of the total 2,574 clients). Reviewing the chart above gives us some indication where to divide each segment using the cumulative percentage line, but let’s take a look at Table 4, which shows the exact cumulative percentage for each poverty likelihood. We see that households with poverty likelihoods between 0.0% and 5.9% account for roughly one-third (34.3%) of the total population. Let’s call this segment “Least likely to be poor.” We also see that poverty likelihoods between 9.7% and 24.8% account for about another third

(31.5%) of the population and that the remaining households account for 34.2% of the population. Let’s call these last two segments “Intermediate likelihood of being poor” and “Most likely to be poor”, respectively.

Let’s take a look at a few characteristics for each segment to see if there are indeed differences in characteristics and behavior among these segments. Running the numbers, we get the following:

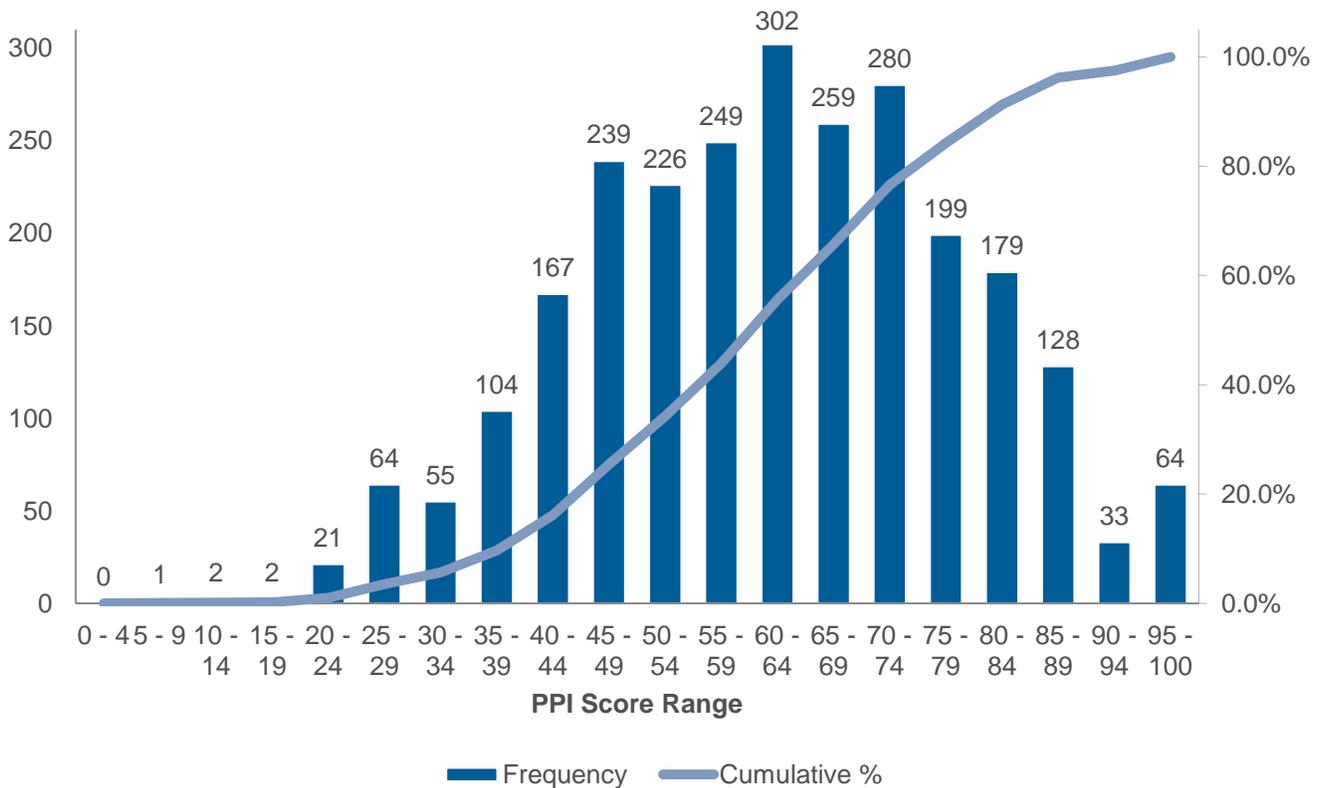
Table 5: Segmenting by Products and Services

Segment	N	Poverty Rate	Age	% Female	# of Products Purchased (Median)
Least likely to be poor	883	3.7%	39	74.7%	9
Intermediate likelihood of being poor	810	17.5%	42	82.3%	5
Most likely to be poor	881	57.3%	47	91.9%	2

The differences among the segments of these clients from the Filipino social enterprise—along with the fact that there are 810 to 881 cases in each segment—are a good indication that the segments are different enough for us to dig further into the data to, for example, understand the type of products most used by those least likely to be poor.

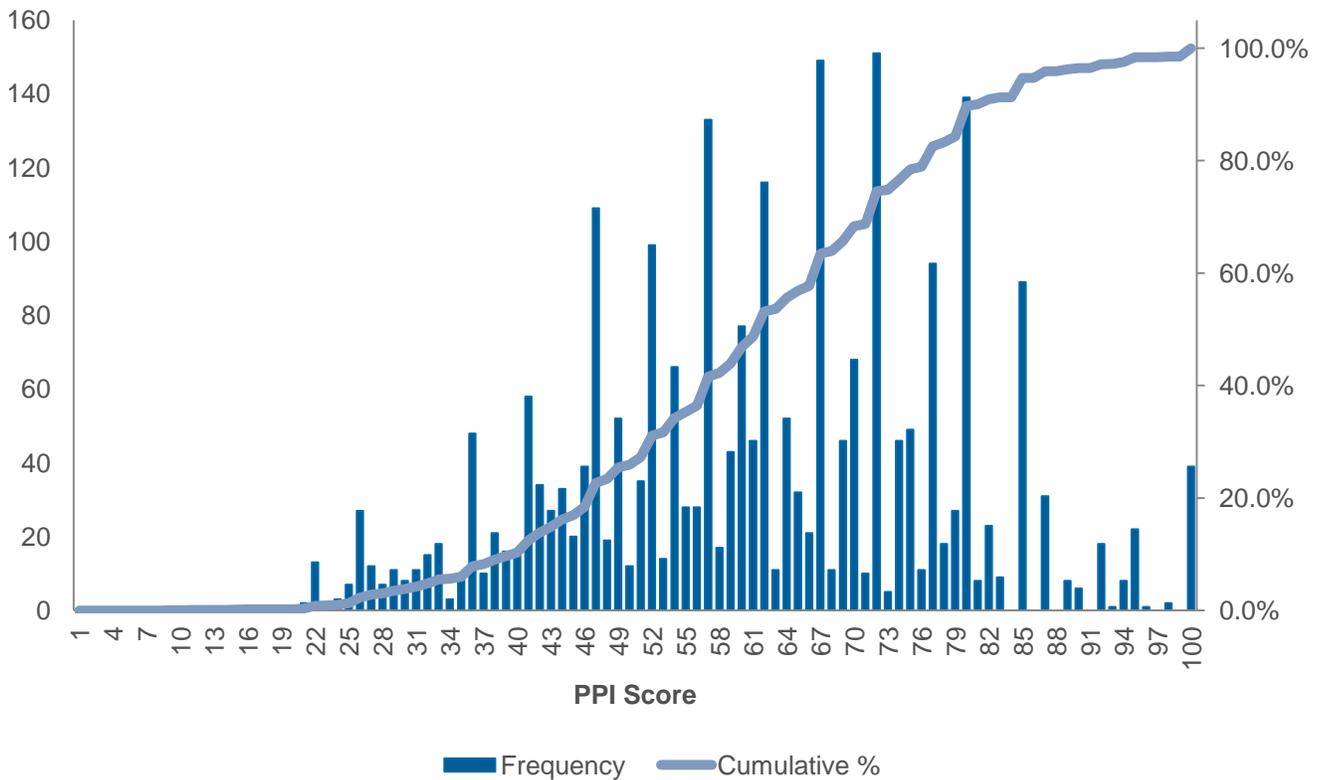
Instead of segmenting by poverty likelihoods, we could have segmented by PPI scores, either by ranges or by individual scores. Segmenting by PPI score ranges would have resulted in the same segments as those above because the score ranges are simply markers for the poverty likelihoods in the look-up tables. Note that the distribution of PPI score ranges in Chart 2 is just the mirror image of the distribution of PPI likelihoods. “Least likely to be poor” becomes “Scores 70 – 100”, “intermediate likelihood of being poor” becomes “Scores 55 – 69” and “most likely to be poor becomes “Scores 0 – 54.”

Chart 2: Segmenting by PPI Score Range



It is also possible to segment by individual PPI scores. Doing so may be more appropriate when segmenting by score range alone does not produce segments that are large enough or comparable in size, if desired. Chart 3 shows the distribution of PPI scores from the Filipino social enterprise.

Chart 3: Segmenting by Individual PPI Scores



Interestingly, breaking this population into near-equal thirds results in the same segments as the preceding two examples (Scores 0 – 54, Scores 55 – 69, and Scores 70 – 100); however, because there are 101 different scores but only 20 different poverty likelihoods and score ranges, cut-offs can often be more refined.

Labeling Segments when using PPI Scores or Poverty Likelihoods

One of the most common mistakes made when segmenting by PPI data is to label the resulting segments in ways that are incorrect and potentially misleading, such as “extremely poor” and “less poor”. These titles are ambiguous and may be interpreted in ways other than that intended. For example, one might wonder, “Why is someone scoring a 54 extremely poor and someone scoring 55 poor?” These titles are also incorrect; PPI scores don’t show you whether someone is extremely poor, or even poor, rather they convey the likelihood of being poor according to a certain poverty line.

Segments should be labeled in a way that identifies how scores have been used (such as “Scores 0 – 14” and “Lowest Scores”) or that demonstrates the likelihood of being poor of such segments (such as “least likely to be poor” or “poverty likelihoods of 35% or more”). The labels assigned to segments influence how the segments are interpreted by readers, so they should be objective and factual.

Segmenting by Poverty Lines

The previous method allowed us to break a given population apart into discrete segments using the PPI but limited our designation of each segment to concepts that are not as readily understood by some readers. The

method in this section allows organizations to discuss their clients in terms of poverty lines, which has obvious advantages in communicating the meaning of each segment. For example, with this method, we can say, “Our beneficiaries who are below the \$1.25/day poverty line have an average daily output of 4.2 rugs.” This would not be possible segmenting by PPI score or poverty likelihood because it is incorrect to say that all clients in the lowest-scoring segment are under the \$1.25/day poverty line.

The disadvantage of this method is the complexity of the analysis, which is less straightforward and requires not only someone who understands precisely how to employ it, but also a system that allows for analysis without undue burden.

The complexity of this approach is due to the inability to separate a population into discrete segments when using it. The PPI (or any other poverty measurement tool for that matter) is unable to discern with certainty whether a household is below or above a poverty line (except for households with 0 or 100% certainty of being above or below such a line). In other words, because we cannot be certain where each household falls relative to poverty lines, we cannot split a population into discrete groups based on them.

To circumvent this issue, we designate a portion of each household to each segment based on the household’s likelihood of falling in each segment. In essence, we are dividing each household into different portions, which correspond to the likelihood of being in a segment defined by poverty lines. When we look at the characteristics of households under the \$1.25/day line, for example, we are just concerning ourselves with the share of these households that are under the \$1.25/day poverty line.

Example

Let’s assume there are four households, each with a probability of falling inside a certain segment, defined by poverty lines, as shown in Table 6. We would like to determine, for each segment, the average age of the heads of household. To do so, we must take the weighted average of all the ages for each segment, with the weights being the likelihood of being in that segment.

Table 6: Example Data

Head of Household	Age	Gender	Likelihood of Living below \$1.25/day	Likelihood of Living between \$1.25/day and \$2.50/day	Likelihood of Living above \$2.50/day
A	80	M	0%	50%	50%
B	60	M	25%	50%	25%
C	40	F	75%	25%	0%
D	20	F	100%	0%	0%

For each segment, add each share of age (calculated as the likelihood that each household is in that segment multiplied by the statistic in question, e.g. age) and divide the sum by the sum of all likelihoods. For example, the weighted average of age, where the weights are the likelihoods that the heads of household are below the \$1.25/day poverty line, is the following:

$$\frac{80 \times 0.00 + 60 \times 0.25 + 40 \times 0.75 + 20 \times 1.00}{0.00 + 0.25 + 0.75 + 1.00} = 32.5$$

Doing this for the remaining two segments allows us to calculate the average age for the remaining two segments, as shown below:

Table 7: Average Age per Segment

Avg. Age of Heads of Household Living under \$1.25/day	Avg. Age of Heads of Household Living between \$1.25/day and \$2.50/day	Avg. Age of Heads of Household Living above \$2.50/day
32.5	64.0	73.3

In this example, we've used a very small population in order to more simply demonstrate this advanced segmenting method; however, with larger sample sizes, we could confidently describe the characteristics of each segment, such as, "Our clients living below the \$1.25/day poverty line have an average age of 32.5 years, which is much younger than clients who are between the \$1.25/day and \$2.50/day poverty lines and those who are above the \$2.50/day poverty line."

The above example uses a continuous variable, but the same approach can be used for categorical variables. For instance, we may be interested in knowing the percentage of clients living between the \$1.25/day and \$2.50/day poverty lines that are female. Using the data in Table 6, we simply translate the values to 1 and 0. Because we are interested in females, we will make female = 1 and male = 0. The same weighted average formula is used as before:

$$\frac{0 \times 0.50 + 0 \times 0.50 + 1 \times 0.25 + 1 \times 0.00}{0.50 + 0.50 + 0.25 + 0.00} = 20.0\%$$

This means that 20.0% of clients living between the \$1.25/day and \$2.50/day poverty lines are female.

When a variable has more than two categories, such as "never married, married, and divorced," create yes/no variables, one for each category, with 0 = No and 1 = Yes, and then apply the process described above.

Conclusion

Segmentation is an important tool we can use when trying to better understand our clients, their behaviors, and the ability of our products and services to meet their needs. While segmentation is powerful, it does not inherently lead to better decisions for the poor. Good data must be collected and managers must ask important questions of that data. The results that segmentation can highlight must be understood and, most importantly, management should act on insights by implementing changes based on them that positively impact the lives of the poor.