Market Analysis at the BoP

PREPARED BY JACOBO MENAJOVSKY AND FRANK BALLARD

Segmentation 101

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.

Companies 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.

In any segmentation exercise, you will analyze several variables that characterize the population in question. Variables that will lead to better information on clients are:

▶ Data on how potential or existing customers are using certain products or services;

▶ Information about customers' needs; and

▶ Demographic data, such as age, geo-location, education level, marital status and poverty status.

You could say that segmentation is part science and part art because it needs to combine a rigorous methodology with a subjective mindset. Ask yourself questions that segmentation can help address before starting any analysis. What is the end goal of the segmentation? What variables are important to include in the analysis?

The following are three important considerations to keep in mind before beginning any segmentation analysis:

▶ All individuals should be grouped into segments, and none of them should be placed into more than one.

▶ Customers pertaining to the same segment should be fairly similar but relatively different from customers in other segments.

▶ Each segment needs to be large enough to be actionable and meaningful for the business.

Typical questions to ask yourself once you have developed your segments are:

▶ Are my segments reachable?

▶ What is the best strategy, product or service that meets their unique needs?

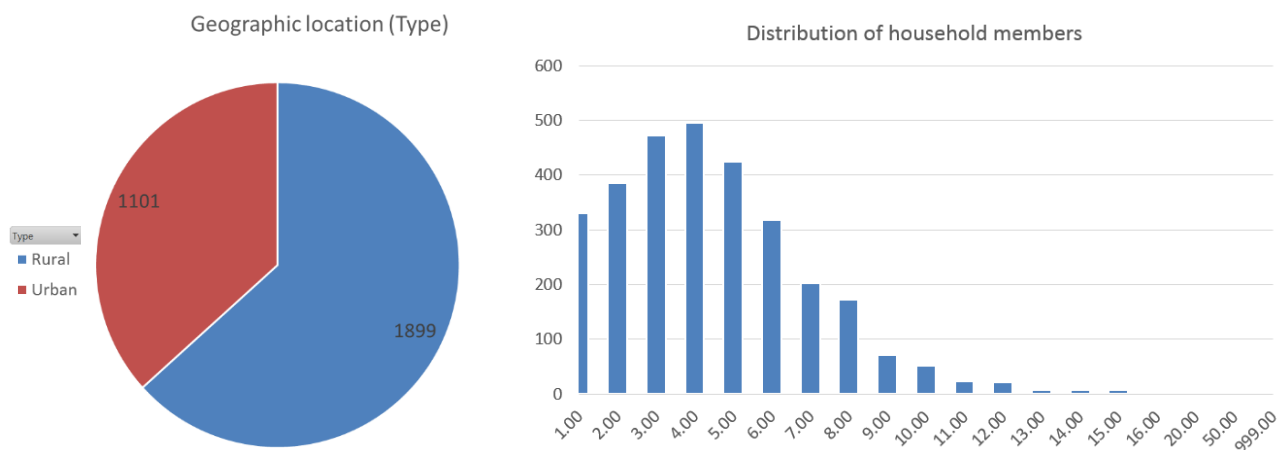▶ What are the goals we are trying to achieve?

Variables

A **variable** is any characteristic, number, or quantity that can be measured or counted and has more than one possible value. A variable may also be called a data item. Age, gender, business income, and country of birth are examples of variables. The value of a variable may vary between data units in a population, and may change in value over time.

There are three main types of variables: nominal, ordinal and numerical. A **nominal** variable is one that has two or more categories, but the categories have no intrinsic ordering. Examples include gender, nationality and possession of a certain asset. Categorical variables that can be ranked or ordered are called **ordinal** variables. Examples include Olympic medals (bronze, silver and gold) or months of the year. A **numerical** variable describes a measurable quantity as a number, such as age or number of children in a household. Numeric variables are quantitative variables.

Distributions

Likely, simply looking at a dataset will not lead to interesting insights. Visualizing the data through charts and graphs allows us to begin asking more interesting questions. The best and easiest way to do this is by checking all variable distributions. Distributions show the frequency of particular values of variables. This is a very simple step that requires minimal data analysis. Let's view the distributions of a few variables using a pivot table in Excel[1]. (Watch **this short video** on building a pivot table in Excel. If you are short on time, skip to the section "Create a pivot table".)

What could you say about these two different variable distributions? Do you see anything strange or unusual?

Checking variable distributions is one of the first steps taken to understand and asses the quality of your data. In our example for the distribution of household members above, the chart indicates that there is at least one case with 999 household members. Don't panic. A convention is to enter 99 or 999 when respondents don't answer a question. Thus, if you wanted to report the average household size, you would want filter those values out or else the average will be skewed.

Dataset

In this section of the course we will be using a demo dataset[2] from a representative sample 3,000 respondents in Kenya. It is included in the document "FII Kenya Wave One Data_Exercise". This dataset is presented in a typical

---

[1] In the video you will watch after reading this document, we will show you how to build these charts. To help with your assignment, create histograms for other variables in the dataset.

[2] The data was collected though a survey by **InterMedia**. InterMedia's program **Financial Inclusion Insights** produces original data and practical knowledge on demand-side trends in mobile money and other digital financial services around the world. FII's research focuses on how people access and use mobile and financial services to improve their lives and livelihoods in eight countries with a combined population of more than two billion: Bangladesh, India, Indonesia, Pakistan, Kenya, Nigeria, Tanzania and Uganda. The demo dataset prepared for this course was made using FII's original dataset, which contains data on 3,000 individuals with more than 1,000 variables. You can access to the original dataset by completing and sending a **request form** to InterMedia.

survey format in which rows represent respondents or units under analysis and columns represent their responses or variables.

Below is a list of each of the variables in the dataset[3]. For each one, review the dataset and write down the type of variable: nominal, ordinal, or numerical.

Header (Variable Name)	Variable Type
Province	
Type	
Duration of the interview	
What year were you born?	
Is the respondent a male or a female?	
What is your marital status?	
Do you currently have a job that earns you income?	
How many members does your household have?	
Do you personally own a mobile phone?	
Do you personally own a bank account?	
PPI. PPI Score	
ppi prob below $2.50 poverty	
AGE.Age scale	
Age.group	
Education	
Adults can access to a mobile device	
Adults have awareness of a mobile money provider	
Ever used MM Adults have ever used MM account	
Adults haven't ever used mobile money	
Adults have ever used a bank account	
PPI_segments	

Go through the demo dataset to create distributions of some of the variables. If you don't know how, skip this and return to it once you have watched the course videos.

Segmentation with the PPI

A common way to measure the poverty of customers is to use the PPI. The PPI allows us to have a clearer idea of the poverty situation of surveyed households, but we can also use it as a tool to segment.

---

[3] We have chosen to use the same variable names applied by InterMedia so that anyone wishing to compare this dataset to the original dataset may do so more easily.

This course will focus on how to segment your market using the poverty likelihoods calculated with the PPI. (While not required for this course, review the **Guide to Segmenting** for a more detailed discussion of this, as well as explanations of using other methods to segment using the PPI.)

The PPI is an excellent example of a variable that can be **transformed** to create a new variable. Using each poverty likelihood as its own segment would result in too many segments, some of which would be too small to be actionable. Instead, we can transform (or convert) the likelihoods into a new variable called *PPI_segments* by categorizing the PPI likelihoods into the segments "most likely to be poor", "least likely to be poor, and another segment for individuals who have an intermediate likelihood of being poor, "mid". Note that while *ppi prob below $2.50 poverty* is a numerical variable, *PPI_segments* is an ordinal variable.

Here's an example of segmenting using the PPI. Remember that the whole dataset contains 3,000 cases. Each has an associated PPI likelihood that the respondent is living below the $2.50/day (PPP) poverty line, called *PPI probability below $2.50 poverty line*. We can order this variable from highest to lowest and separate the client population into three segments of equal size. The first 1,000 are categorized as the "most likely to live below the $2.50 line", the following 1,000 are categorized as "mid", and the last 1,000 are categorized as "least likely to live below $2.50". Below, we obtain the poverty rate for each of these three segments.

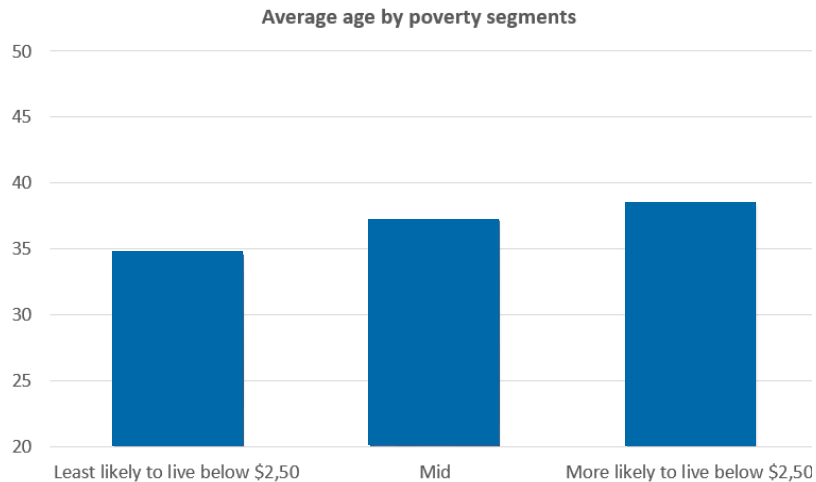| PPI Segments	$2.50/day (PPP) Poverty Rate
Least likely to live below $2.50	11.7%
Mid	47.8%
Most likely to live below $2.50	89.3%

This means that among those who were grouped as the least likely to live below the poverty line, only 11.7% actually live below the poverty line. On the other end, almost all (89.3%) of the respondents grouped as the most likely to live below the poverty line do in fact live below the poverty line. Finally, the category "Mid" fall in the middle of our ranking and 47.8% of the respondents in this category live below the poverty line.

We chose to use three equal segments to provide a simple introduction to segmentation. We could have segmented the population with the poverty likelihood in countless ways. For example, we can split the population into five equal segments or simply say that anyone whose likelihood of being poor is below 50% is placed in one segment while all others are placed into another. Just remember that segments should be large enough to be actionable and different enough from one another to be meaningful. You can refer to the **Guide to Segmenting** if you would like to see other examples.

Analyzing our Segments

We now know that these three segments are significantly different from each other in terms of poverty, but are they different in other ways? To answer this question we turn again to the data and run some analysis. As we look at how the segments compare using other variables, we can build a profile, or **customer persona**, of our three poverty segments.

Most likely we will find that some variables really differ from segment to segment while others don't. Let's check how the age of the respondent compares. Instead of observing the frequency of each age in each segment, we can use the pivot table to quickly calculate the average age of each segment. The chart below shows the average age for the segments.

Average age by poverty segments

What could you say about age differences when comparing poverty segments? The answer is subjective and depends on the business question at hand. In this case, we might recommend that if a business is developing a communication strategy for potential customers, it should be aware that the average respondent across all segments is in their 30s, with the BoP segment being slightly older on average.

Introducing our Business Challenge

All-Bank, a fictitious commercial bank in Kenya that provides mobile financial services in addition to traditional banking services, wants to increase their market share and penetration, but in doing so, they want to serve higher numbers of the individuals situated at the BoP that lack bank accounts. The demo dataset we have been using is the same dataset they have received.

After segmenting and analyzing the available market data, we will be better prepared to understand the BoP segment and craft a set of strategies and actions to successfully reach the BoP.

We already segmented this dataset using the PPI. Now let's investigate potential differences between the poverty segments. We should again observe the data that is available and choose what information best relates to the business challenge.

From all the available variables in this dataset, we have chosen to use the following:

- ► Type (Rural-Urban setting)
- ► Education
- ► How many members does your household have?
- ► Do you personally own a bank account?
- ► Do you personally own a mobile phone?
- ► Adults can access to a mobile device
- ► Adults have awareness of a mobile money provider
- ► What is your marital status?

What differences between poverty segments exist across these variables? After spending some time analyzing the data, you should start discovering certain patterns and differences among our three segments. Discovering

which variables differ *and* which don't add a lot of information to our analysis and will serve as our groundwork in building differentiated strategies and actions for various segments.

How would you help them to characterize potential new customers at the BoP, which we grouped under the "most likely to live below $2.50/day" category?

Now watch and follow along with two short videos on how to analyze these variables.