Demystifying Segmentation – Part 2

One objective in my last article on segmentation was to at least convey the notion that segmentation can mean different things to different people. In one case, it may simply be using business rules such as select all customers who have been a customer for longer than 2 years, live in Quebec and have bought more than 2 products. In another case, it may be scoring customers through a model or perhaps a value-type metric and then selecting these customers based on score. Another scenario may involve the use of statistics to determine distinct or homogenous groups of customers. No one approach is necessarily superior to the other. They are all good approaches as each one may be more appropriate depending on the particular business problem and more importantly the data. It was indicated that both simple and complex solutions can be acceptable to a particular business stakeholder. Not all segmentation requires sophisticated statistics. In some cases, a simple and more pragmatic approach will suffice. But the question remains on when is it appropriate to adopt the simple versus complex approach.

Simple vs. Complex Approach

The first consideration depends of course on the data environment. Why? More data implies that we can do more things with it and ultimately use it for more marketing-related activities. But this will depend on both the quality as well as the variety of data. The more complex solutions will arise from these types of data-rich environments.

The second consideration depends on the volume of customers. A larger volume of customers offers more potential to derive significantly larger dollar benefits even when lift becomes more marginal. Consider the situation of a 1% business lift on 2 million customers versus a 5% lift on 200,000 customers. From an absolute standpoint, the larger opportunity is with the larger volume of customers despite its much lower lift potential. With this larger volume of customers, more complex type solutions can be explored.

Let us take a look at an example to bring some of these scenarios to life. Suppose a company sells one product and only collects the billing information of the customer for the last 2 years as well as name and address. The current number of customers is 100,000. Before doing anything, though, we need to first ask ourselves how this information is going to be utilized. More specifically, is this information going to be used for targetting or for communication purposes. Of course, most people would say both. Yet, in reality, a specific solution will only be able to

resolve one of the above purposes. For instance, if targetting is our prime objective, we certainly don't need clustering. One could merely rank order names based on the desired business objective, which in this case is billing amount or sales. Through rank-ordering, customers are placed into groups or deciles where the top decile represents the highest billed customers and the bottom decile contains the lowest billed customers. Groups of customers can then be selected for a marketing initiative based solely on their prior total sales (billing) with the company. However, what if the marketer wants to establish more meaningful communication with this company? The question to ask here is what other information besides sales can be used to help in this process. In the example here, there is no opportunity to use other information for communication purposes since no other information beyond billing amount exist.

However, suppose this same company collects payment type and keeps track of a given customer's billing history. Now, we are able to create two other key variables. The first one, tenure, is developed using the customer's first billed date as a proxy for tenure. Meanwhile, payment type can be used to separate customers into pre-authorized payment versus non pre-authorized payment. Along with this other newfound information, we may have identified our target group of customers as being the top 50% based on their total customer billings in the last year. We may now ask ourselves whether or not this tenure and payment type information is rich enough to develop unique communication programs. Indeed, a simple clustering exercise might help to answer this question in a quantitative manner by allowing us to scientifically determine whether there are some distinct customer groups based on tenure and pre-authorization plan to these top billers. The data and results might look as follows:

Cluster 1	Average
Tenure	2.5 yrs.
% with Pre-authorization plan	40%

Cluster 2	Average
Tenure	6 years
% with Pre-authorization plan	20%

From the above, the clustering exercise does provide additional insights if we can somehow develop a unique communication program to newer customers who are more likely to pay by automatic withdrawal versus a communication program to longer-tenured customers who pay by cheque or credit card.

Profiling vs. Clustering

Yet, besides the clustering approach, another option might be to profile the top billers (top 50%) vs. the bottom billers (lower 50%). The difference with profiling versus clustering is that the profiling approach utilizes an objective function which is whether or not the customer is in the top 50% of customer billers. Other customer characteristics are then analyzed in order to identify which are the key characteristics in differentiating a top customer biller from the bottom 50%. We call this approach supervised learning as the objective function of being in the top 50% or not determines (supervises) what are these key differentiating characteristics.

In the clustering approach, there is no one variable which determines the outcome of the other variables. As stated in the previous article, the objective of clustering is to identify those characteristics which best assign customers into unique and distinct groups. Our objective here is not to optimize a specific variable or metric but rather a scenario where variation between

customer groups or clusters is maximized and variation within customer groups or clusters is minimized. The data or variables in this case are "unsupervised" or not analyzed against a specific customer variable or metric.

Looking at the Complex Example

In our example, here, we can then actually determine if either tenure or preauthorized payment plans have an impact on being a top biller. If either or both of these characteristics have an impact on being a top biller, then communication programs could be built around the top 50% without doing any clustering. Yet, if profiling is the more reasonable and practical option in this case, we may ask ourselves why we don't simply segment all customers based on value and then simply profile rather than cluster the high value customers from regular customers. The answer in many cases is that this is an acceptable solution. But in cases of companies with large customer volumes (i.e., much larger than 100,000 customers) and very rich data, this may not be the optimal solution. For example, a given bank may have 5 million customers where we deem that the top 2.2 million customers have enough value (high value) to be considered as our net eligible group for future CRM initiatives. Profiling these top 2.5 million against the bottom 2.5 million reveals six key characteristics that best differentiate both groups:

- High tenure
- Live in Toronto
- Have a mortgage
- Have multi investments
- Have multi loans
- Have Visa
- Have high transaction fees

You may surmise from the above findings that this information falls into the "so-what" category by rightfully arguing that you could have created this profile of a high value customer without doing any data mining. So the question remains about how we use this information above to develop unique communication programs. Remember, we already know our target group is 2.5 million customers. But how do we talk to them? This is where clustering plays a hugely significant role. Through clustering, the analysis could provide these types of clusters:

Cluster 1	Average
High tenure	3
Live in Toronto	**0.8**
Have a mortgage	**0.9**
Have multi investments	0.3
Have multi loans	0.4
Have Visa	0.6
Have high transaction fees	0.5

Cluster 2	Average
High tenure	**10**
Live in Toronto	0.5
Have a mortgage	0.4
Have multi investments	0.35
Have multi loans	0.5
Have Visa	**0.9**
Have high transaction fees	**0.8**

Cluster 3	Average
High tenure	3.5
Live in Toronto	0.45
Have a mortgage	0.49
Have multi investments	**0.7**
Have multi loans	**0.8**
Have Visa	0.65
Have high transaction fees	0.55

As you can see from the above information, three separate programs could be developed to high value customers. For Cluster 1, programs would be developed that perhaps communicate the advantages of different type of long-term lending schemes (mortgages) as well as particular financial benefits that are more advantageous to someone living in a large city like Toronto. For Cluster 2, programs might be developed that speak to providing benefits to credit card users as

well as benefits to high transaction users. Meanwhile, Cluster 3 programs could be developed for the type of customer who is more financially sophisticated both in terms of investments as well as loans.

This type of approach allows marketers to use the information in a two-pronged manner which is to both target the best customers based on a defined metric but also use the other information to meaningfully communicate to them.

Richard Boire is a principal partner at the Boire Filler Group, a data mining consulting company.