Survey of Behavioral Segmentation Methods

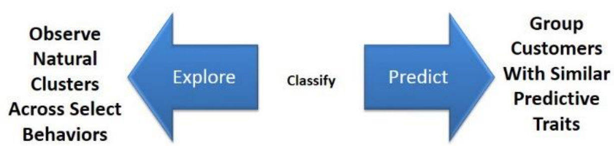
Written by
Rhonda Carraway Petty
Marketing Insights Data Scientist
rpetty@mathnetix.com

There is no magic rule or formula for deciding how to classify customers into a smaller number of homogeneous groups, we call 'segments'. But, there are rules of thought that set the Marketing Scientist on the right path.

This paper presents a practical synopsis of different approaches to behavior-based customer segmentation. It is written, not as an exhaustive academic disclosure of all possible analysis paths one can pursue, but to guide the development thought process. It is written, not as a mathematical explanation, but in simple-to-understand terms for business executives, junior analysts and others who are considering ways to segment their customer population and simply need a place to start. Occasionally, the writing flows in the first person to share empirically-derived opinions.

STATE THE OBJECTIVE

The first question the business executive and Marketing Scientist must answer (together) is 'What's the point?' In what ways will the segments be used? Is the purpose to explore behavior or predict behavior? Often, it's both---explore behavior so you can predict behavior. The distinction is important because it helps the scientist select the most applicable statistical method for classifying customers into mutually exclusive homogeneous groups.



The statistical method used to create segments is all but dictated by the objective of the exercise.

Exploratory: If the objective is *exploratory*, then the 'kitchen sink' approach can be an insightful first step. In this scenario, an array of customer behaviors and attributes, many correlated with one another, are tossed into the 'kitchen sink'. Through statistical analysis enabled by popular software packages, customers similar across the spectrum of behaviors are grouped together. Although straight-forward in concept, the Marketing Scientist manages the statistical processing of the data, observes results and experiments through iterative analysis to find the 'optimal' number and fit of the segments.

Predictive: If the objective is *predictive*, then identifying the behavior to predict is the first step. Select a future behavior that, if known, will be used to make smarter marketing decisions. Behaviors like promotion response, repeat and cross-sell are popular subjects for predictive segmentation schemes. If, for example, the Marketer knows which existing customers are likely to repeat, he can take action to motivate the repeat---to help move it along---varying his strategy based on the repeat 'grade' or segment.

CREATE THE ANALYTICAL DATA FILE

After defining the segmentation objective, the Marketing Scientist must create an Analytical Data File. This file is the data feed into the statistical process that creates the segments. The Analytical Data File contains one row for every unique customer who is typically identified with a unique customer ID. The customer's behaviors are numeric variables represented as columns in the data file.

The numeric variables (columns) are derived by the Marketing Scientist from transactional details. The Marketing Scientist summarizes the transactional details to create behavioral variables that form the basis of the segmentation logic. Current behaviors believed to correlate with the customer's future value are usually the behaviors that feed the segmentation routine---variables like number of transactions, spend per transaction, elapsed days since the last transaction and yes/no permission flags are a few examples of such variables.

Figure 1 illustrates the arrangement of an Analytical Data File given only five customers and four behaviors. In reality, the Analytical Data File is much deeper and wider. If there are several million customers, then the Analytical Data File will be several million rows deep. If there are several hundred behaviors, then the Analytical Data File will be several hundred columns wide.

Figure 1. Structure of Analytical Data File

Customer ID	**Behavior 1** Number of Transactions	**Behavior 2** Spend per Transaction	**Behavior 3** Elapsed Days	**Behavior 4** Signed up for Email Alerts
1	5	59.82	15	1
2	3	31.17	45	0
3	1	74.19	105	0
4	3	42.30	92	1
5	4	64.20	78	1

SELECTING THE SCIENCE

A brief description of popular statistical routines used to create behavior-based customer segments are presented in a loosely ordered fashion from exploratory to predictive. The methods presented are not an exhaustive list of techniques, but do introduce the reader to methods commonly used. Readers interested in technical details of any one approach should seek counsel from a good book in statistics.

Principal Component Analysis (PCA). Combines a large number of potentially correlated behaviors into a smaller number of uncorrelated principal components. This technique is useful when working with a large number of behaviors that need to be reduced into a more manageable number of behaviors before creating the final segments. Rather than using intuition to decide which behaviors to keep, Principal Component Analysis 1) retains all behaviors by transforming them into uncorrelated principal components or 2) explores behavioral relationships to guide decisions related to which variables to keep and which variables to drop---dropping those that have little added value because they are strongly correlated with other behaviors.

Principal Component Analysis is used to:

- Reveal major underlying behavioral 'phenomena'
- Reduce a large number of behavioral variables into a smaller number of core components

> **Case Study**: After deriving a large number of variables to describe the banking behavior of checking account customers, PCA was used as a data-reduction technique. Through PCA analysis, correlated behavioral variables were combined into a smaller number of uncorrelated core behaviors. Then, those core behaviors for each customer were passed through statistical clustering to create the final segments.

Statistical Clustering. Groups customers with similar values for multiple variables. Customers most alike across the set of variables, taken as a whole, are placed into the same segment. Customers in the same segment share similar values, not exactly the same values, across the set of variables. Each segment is called a cluster and each cluster is described with a numeric 'centroid'. The numeric centroid describes the average behavior of the customers within the segment. The centroid is the numeric 'face of the segment'.

Following the creation of initial clusters by the statistical software (under the control of the Marketing Scientist), each customer is assigned to her closest cluster based on its Euclidean distance to each centroid. In an iterative fashion, customers are assigned to their closest cluster, centroids are recalculated and customers are considered for re-assignment. This process continues until a defined stopping condition is satisfied. Clustering is computationally intense because of its many iterations. However, many software packages have embedded 'quick clustering' algorithms.

Statistical Clustering is used to:

- Group customers who share similar values across a set of behaviors
- Identify outliers (customers in small segments)

> **Case Study**: After creating uncorrelated core components using PCA, statistical clustering was used to group customers sharing similar values across the set of components taken as a whole. Iterative runs of the clustering process with visual inspection and statistical measures of 'fit' guided the optimal number of segments.

Classification Tables. Groups customers who share the exact same values for a small number of distinct behaviors, usually no more than two or three behaviors. The Marketing Scientist has to decide which behaviors will define the groups. Typically, behaviors correlated with a customer's future value or propensity to repeat are selected. Or, instead of behaviors, attributes like age and gender might be selected, depending on how the segments will be used:

Classification Tables are used to:

- Place customers into temporal groups to satisfy a specific and often immediate need
- Devise if-then-else marketing tactics that vary by customer segment

> **Case Study**: A company wanted to understand the impact of a marketing promotion by age and gender to confirm their hypothesis that younger men are harder to incite. Customers were placed into a grid that represented the cross of age and gender. Then, the impact of the promotion was measured on an overall basis as well as for each age-gender segment.

R-F-M. This widely accepted classification approach combines customers who share the same **R**ecency, **F**requency and **M**onetary Value and, as such, are expected to share the same propensity to repeat. Because these variables are often strongly correlated, the Marketing Scientist should re-express them into less correlated variables. In this way, each of the three variables will offer a unique perspective about the customer's behavior. For example, Monetary Value depicted by Total Spend is highly correlated with purchase Frequency (the more times one purchases, the higher their cumulative Total Spend). But, Monetary Value depicted by Spend-per-Transaction is less correlated with purchase Frequency (Spend-per-Transaction may be influenced by the number of transactions, but less so than cumulative Total Spend).

R-F-M is used to:

- Group customers who share the same Recency, Frequency and Monetary Value and, thus, the same inferred propensity to repeat
- Devise if-then-else marketing tactics that vary by customer segment

> **Case Study**: After observing the respective R-F-M distributions of the customer population, each individual customer was mapped to one of three(3) Recency groups (high, medium or low), one of three(3) Frequency groups and one of three(3) Monetary Value groups. Crossing the possibilities yielded 3x3x3=27 segments. These 27 segments were reduced into a smaller number of more manageable groups using customized business rules. Finally, a profile of each segment was created to give the Marketer more insight into the 'personality' of each group.

CHAID. CHAID stands for Chi-Square Interaction Detector. It is a statistical process used to explore the relationship of variables to a specific variable of interest—a variable the Marketer wants to predict. CHAID is visualized as a decision tree that begins with all customers in the same segment. P% of the customer population in the single parent segment exhibits the desired behavior. At each level of the tree, customers are separated, conditioned on their value for another variable, into smaller segments having a higher and higher concentration of customers exhibiting (or not) the desired behavior. CHAID produces 'pockets' of customers with a high density of the behavior of interest, assembled from the values of other 'predictive' variables. These 'pockets' are segments with an above-average or below-average propensity to exhibit the behavior of interest.

CHAID works by calculating lots of cross-tabs behind the scenes to enable 'tests of independence' between the parent node's variable (the behavior to be predicted) and the other variables (the behaviors that predict it). The parent node's variable is "dependent" on another variable if there is a strong association between them. That variable to which the parent variable is most dependent is used to split the parent node into smaller groups. This process continues in a tree-like fashion until a stopping condition is satisfied.

CHAID is used to:

- Create if-then-else rules for predicting behavior
- Discover the relationships among variables
- Suggest ways to combine different levels of a variable into a new variable
- Place customers into segments that share the propensity to behave in a specific way

> **Case Study**: After launching a catalog campaign to acquire new customers from purchased prospect lists, a retailer wanted to know what list attributes best differentiated responders. This knowledge would help the retailer select good lists in the future. To support the retailer, a sample of catalog responders and non-responders were selected, stratified across the different purchase lists. With variables that described the list from which each prospect was obtained, CHAID was used to find the list attributes most predictive of acquisition. Now, the Marketer knew the particular list attributes to pursue in the future when considering the purchase of names for acquisition.

Regression. Regression marries multiple variables to produce an equation for predicting a pre-specified behavior. The behavior to predict is called the 'dependent variable'. The variables that predict it are called the 'independent variables'. Regression uses statistical principles to find the subset of candidate independent variables---and their respective influence---that best predict the dependent variable. Linear

Regression is a forecasting technique used to predict a continuous outcome---like sales in dollars. Logistic Regression is a forecasting technique used to predict a binary outcome---likes yes/no, the customer will respond.

Regression is used to:

- Produce business forecasts that can be explained by other known things
- Predict the propensity of a customer to behave in a specific way

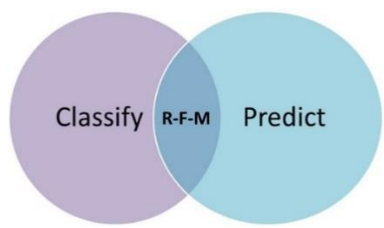
Case Study: A Marketer wanted to create a retention strategy that varied for each customer. He wanted to vary the strategy and offer richness based on the customer's propensity to repeat. A Logistic Regression model was created and used to rank customers from high-to-low propensity to repeat. The Marketer implemented a contact and retention strategy that varied for each segment.

R-F-M Variations

R-F-M walks the line between 'classification' and 'prediction', making it an attractive method for creating initial customer segments. It offers the simplicity of classification with the advantage of prediction. R-F-M segments are assembled in a manner resembling a cross-tabulation while using behaviors predictive in nature. Because R-F-M nests between classification and prediction, its underlying processing logic will tend toward Classification or it will tend toward Prediction:

- **RFM Classification**. Segments consist of customers who look alike on each of the R-F-M dimensions
- **RFM Prediction**. Segments consist of customers who look alike on an R-F-M numeric score

Figure 2. R-F-M Nests between Classification & Prediction



Classification R-F-M

The traditional approach to R-F-M is simple computationally and conceptually. It produces groups of customers who are exactly alike on each R-F-M dimension, operating like a three-way cross-tab. When creating R-F-M segments using a classification approach, there are three primary steps:

- **Summarize Behavior**. Summarize historical transactions into variables that depict Recency, Frequency and Monetary Value

- **Establish Break-Points**. Observe the distribution of these variables across the customer population to select high-medium-low break-points for each of the three dimensions

- **Cross and Reduce**. Cross and combine the high-medium-low rankings of the three dimensions to create a manageable number of final segments, typically on the order of 8 to 12 groups

Figure 3. Classification RFM

The goal of the Marketer is to move customers to higher-performing R-F-M segments

Prediction R-F-M

Another approach to R-F-M operates similar to a predictive model. Instead of producing a three-way cross-tab, the prediction-based R-F-M approach calculates a numeric score for each customer based on the continuous R-F-M metrics. The R-F-M scoring equation can be generated using regression analysis or it can be intuitively derived based on the following presumptions about the relationship of R-F-M with a customer's propensity to repeat:

- Recency: As recency increases (meaning, the last trip was <u>not</u> recent), repeat decreases
- Frequency: As frequency increases, repeat increases
- Monetary Value: As monetary value increases, repeat increases

With these inferred relationships, an intuitive R-F-M Score is derived to force a higher score when Frequency and Monetary Value increase---and a lower score when Recency increases.

The equation of Figure 4 depicts an intuitively derived scoring equation. In this example, the R-F-M score simplifies into the customer's cumulative Total Spend weighted up or down depending on how recent the customer last purchased.

Figure 4. R-F-M Scoring Equation Example

R-F-M Score = (Frequency * Spend-per-Trip) * 1/Recency

R-F-M Score = Total Spend * 1/Recency

The R-F-M scoring model is a simple and useful way to decide who to include in a marketing promotion. The Marketer simply sorts the customer list by RFM score. High scores depict a high propensity to respond and low scores depict a low propensity to respond. The R-F-M scoring model is a quick way to sort a customer file based on predictive traits. However, it is theoretically possible for two customers to receive the same R-F-M score even though their individual behaviors that feed the score are different. Thus, if the Marketer needs to vary specific things like message, creative or other tactic based on component behaviors, the R-F-M classification approach may be preferred. The R-F-M classification approach preserves insight into the customer's specific Recency, Frequency and Monetary Value behavior.

Hint: Consider using both techniques. Create discrete groups of segments using the R-F-M classification approach. Then, use the R-F-M score to sort customers within each discrete group as a way to establish marketing priorities *within* each segment.

Concluding Remarks

As an analytical consultant specializing in database marketing strategies, I have designed behavioral segmentation schemes across many industries. The blueprints for each are custom designs that consider each company's unique business. Some applications have been for companies with shallow diversity in product and price. Other applications have been for companies with multiple lines of very different mini-businesses. Through these experiences, I have cultivated opinions and rules of thought that shape the development roadmap with each new segmentation assignment. What follows are helpful hints intended for Marketers who want to build their own customer segmentation schemes.

Hint #1: Think about the behavior you want to change.

If it is behavior you want to change, then observe behavior and how it is changing. Most Marketers want their customers to be more loyal. They want to improve loyalty. But, what exactly is loyalty? If you could describe it with one variable, what would it be? Would it be frequency, value, consistency? Decide. Then, get a good handle on that one attribute before folding in other variables. This will keep your approach simple and straight-forward before diving in with added complexity. If your design is too complex, there is greater risk the segments will not be used or used correctly.

Hint #2: Use root variables.

Dividing customers into groups of high, medium and low value is a popular segmentation scheme, but how actionable is a strategy based only on a customer's past value or a customer's expected future value? Isn't it more actionable to understand the behavior that produced the value? Two customers might have the same value although very different in their root behavior. Segmentation is more actionable if it doesn't lose sight of the core behavioral variables that correlate with the customer's value.

Hint #3: Avoid variables not under the direct control of the customer.

Hold each variable you are considering as a core dimension in your segmentation scheme up to the 'Can the Customer Control It?' lamp. Avoid variables not under the control of the customer. For example, although age and gender may be predictive of future behavior and worthy as an input into regression models, they are not behaviors in and of themselves. Even changes to variables like age and gender are not behaviors. Everyone has a birthday. Don't anchor your behavioral segmentation on these types of descriptive attributes. Instead, use descriptive attributes to build predictive regression models, passing the regression score to your segmentation model as you advance your design.

Hint #4: Include a measure of migration.

Marketing strategy is not 'High', 'Medium' and 'Low'. Instead, marketing strategy is what you will do---things like 'Thank', 'Reward', 'Win-Back'. When you incorporate behavioral *change* into your segmentation process, you force the output to be groups of customers requiring the same marketing action. For example, if a customer is *consistently* loyal with high spend and high frequency, "Thank" this customer. If a customer who was once viewed as loyal is now transacting less, a "Win-Back" strategy may be the appropriate action. Bake 'change' into the segmentation scheme so the segments in and of themselves guide action.

Hint #5: Split on tenure first.

R-F-M are variables summarized from transactional details over a fixed duration of time. Before customers can be compared on their values for R-F-M, those values must have been calculated over the same duration—otherwise, comparisons are 'unfair'. Because new customers have not 'been around' as long as tenured customers, R-F-M summarized from the last 12 months, for example, will be incomplete for new customers. Pivot your approach on tenure first---then create the R-F-M segments.

Hint #6: Use a rolling time horizon.

Base your segmentation on behaviors observed in a fixed interval of time that you advance on a rolling time horizon. This enables you to define relatively static high, medium and low break-points. If behavior is cumulative over time, then their numeric values get higher and higher---and the break-points for high-medium-low also have to get higher and higher to preserve a proportionate mix of customers by segment. In other words, don't base your design on cumulative behavior. Base it on behavior within a fixed time window---then roll the time window forward with each segmentation update.

Hint #7: Watch out for seasonality.

If your approach includes a component of behavioral change, then watch out for seasonality. Make sure changes over time depict an exceptional shift in the customer's behavior and not a natural shift because of the season. Use historical distributions to establish a tolerance for what is labeled 'exceptional'.

Hint #8: Plan for updates.

Design your segmentation logic with updates in mind. Select an approach that is easy to implement and maintain. Feed your design with variables that can be obtained easily and for everyone. Select an update frequency that aligns with customer patterns at large. For example, businesses marked by customers who transact often (banks, grocers…) will benefit from a frequent update schedule. Businesses depicted by customers who transact less often (automobile dealerships, luxury travel….) need a less frequent update schedule. As a general rule of thumb, update frequency should align with the average purchase cycle.

mathnetix®
just in time customer connections®

Rhonda Carraway Petty
rpetty@mathnetix.com
601-384-2488

Rhonda is a Marketing Insights Data Scientist with a proven track record in crafting highly actionable segmentation schemes based on behavior and predictive insights. Throughout her career experience, she has designed segmentation solutions for several large national brands in the financial services, airline, consumer durables, grocery and retail sectors.