SAS6222-2016

Enhanced Segmentation Using SAS® Visual Analytics and SAS® Visual Statistics

Darius Baer, Ph.D., and Suneel Grover, MBA, MS, SAS Institute Inc.

ABSTRACT

What will your customer do next? Customers behave differently; they are not all average. Segmenting your customers into different groups enables you to provide relevant communications and interactions, resulting in greater customer satisfaction as well as increased profits. SAS® Visual Analytics and SAS® Visual Statistics enable you to visually analyze your marketable universe to create more useful segments for customer relationship management and to understand the value and relative importance of different customer attributes. You can segment your customers by using the following methods:

1) Business rules
2) Quantile membership
3) Supervised classification
4) Unsupervised clustering

Whichever way you choose, SAS enables you to graphically represent your customer data with respect to demographic, geographic, and customer behavioral dimensions. This paper covers the four segmentation techniques and demonstrates how SAS Visual Analytics and SAS Visual Statistics can be used for easy and comprehensive understanding of your customers.

INTRODUCTION

How can you maximize relevance with your customer and, in doing so, increase your profits? Because customers have different reasons for interacting with your brand, they behave differently—they cannot be lumped into one group. In the age of the customer, anticipating their needs, meeting their expectations, and addressing their pain points is critical. Customers are more empowered and connected than ever before, with access to information anywhere, anytime—where to shop, what to buy, and how much to pay. Brands realize it is increasingly important to predict how customers will behave and to respond accordingly. Because consumer behavior varies, data-driven marketers can use segmentation methods to group them by common needs and attributes. Simply put, the deeper you understand your customers' buying habits and lifestyle preferences, the more accurate your predictions of future buying behaviors will be.

You need to treat your customers differently depending on the segment to which they belong. By providing personalized offers and communication to each segment, you can do the following:

1. Increase customer lifetime value
2. Reduce irrelevant customer interactions
3. Differentiate your brand relative to your competitors
4. Generate higher profits

In the era of ever-expanding data and hyper-personalization, is segmentation outdated or more effective than ever? That is contingent on your approach and methodology, but many data-driven marketers continue to prioritize segmentation as the foundation of customer insights across their companies. A variety of segmentation strategies exists to stimulate customer preferences for a brand and to expand the ability to understand unique behaviors of consumer expectations—applicable to both current and prospective customers.

Marketers have used segmentation as a technique to target audiences for communications, products, and services since the introduction of customer relationship management (that is, CRM) and database marketing. Within the context of Customer Segmentation Intelligence (CSI) (Baer, 2012), there is a variety of attributes, ranging from consumer demographics, geography, behavior, psychographics, events and cultural backgrounds. Over time, segmentation has proven its value across every imaginable industry, and brands continue to use this strategy throughout every stage of the customer journey:

1. Acquisition
2. Upsell/cross-sell
3. Retention
4. Win back

Segmentation is a proven marketing technique with a number of use cases and benefits, but does analytic sophistication matter? According to Forrester Research in their Forrester Wave™: Customer Analytics Solutions, Q1 2016 report authored by Brandon Purcell, no matter where "customer insight professionals sit within the marketing organization, line of business, or centralized customer intelligence team, it is their job to use advanced customer analytics to generate insights."

The report proceeds to state that "often data scientists with profiles in advanced analytical algorithms and programming languages such as R and Python are more adept at wrestling with mountains of messy customer data. Since the demand for these data scientists far exceeds supply, this job is increasingly the responsibility of marketers who need marketer-friendly tools for data management and analytics."

But do you need Ph.D. talent to advance your customer analytic capabilities? We are living through an exciting era of innovation in approachable analytic technology. The last important takeaway that Brandon Purcell shares in his report is the incorrect assumption of using traditional analytics to uncover deep customer insights. [2]

Calling for the usage of a new breed of algorithms, methods, and advanced analytical techniques, leveraging all your data, and moving past the reliance on reports and dashboards, there is always room for improvement—no matter where your organization is in the usage of segmentation.

Now, let's reflect on a few questions:

- How many marketers do you know who have no issues with data management?
- How many marketers do you know who have too many data scientists on their teams?
- How many marketers do you know who have the time to manage multi-channel interactions between their marketable consumer audiences and their brand, while taking the time to drill into the details of k-means clustering, decisions trees, and other forms of advanced algorithms?

It is our belief that data-driven marketers need to be enabled by approachable and actionable advanced analytics to make more powerful decisions within today's complex and interconnected business environments. The big picture boils down to one, two, or three core enablers, based on your organization's goals and preferences:

Figure 1: Analytic Technology Enablers

SAS Visual Analytics and SAS Visual Statistics represent a category of interactive and collaborative technology to provide a path for the marketer to be curious and innovative. The tools represent an intersection of data management, interactive visualization, and big data analytics joining forces in ways that we have never seen before, while lowering the entry threshold to exploit sophisticated data science for segmentation.

SAS Visual Analytics and SAS Visual Statistics empower you to visually group your customers based on common attributes. Your purchase history for each customer should enable you to see the affinity that different customer groups have for different products as well as to associate those affinities with various demographic characteristics. You can perform segmentation using business rules or get more innovative using advanced analytical methods such as supervised decision trees or unsupervised clustering.

This paper will demonstrate the capabilities of using SAS Visual Analytics and SAS Visual Statistics to segment your customer population in ways that enable you to more easily interact with and manage your customer relationships. Using SAS Visual Analytics and SAS Visual Statistics provides all the functionality needed for segmentation in an easy-to-use and very fast interface without having to write or manage computer code. By using this point-and-click interface, you can quickly focus in on your marketing strategies, extract the needed information, and move forward managing your customers.

There are three expectations that a user should have from a computer application:

- Function
- Performance
- Ease-of-use

SAS Visual Analytics and SAS Visual Statistics have all three. In this paper, we will focus on the function and ease-of-use of SAS Visual Analytics and SAS Visual Statistics SAS VA and VS for creating segments. Moving forward, we will first discuss the why and what of segmentation. Then we will discuss the how, specifically referencing SAS Visual Analytics and SAS Visual Statistics for visually creating, understanding, and using your segments.

MAKING THE BUSINESS CASE FOR SEGMENTATION

If you treat your entire marketable universe the same through offers and communication, you will only satisfy a small portion who purchase, resulting in smaller profits, irrelevant interactions, and subpar brand experiences. We are currently living in the age of big data, hyper-connectivity, and dynamic personalization, and segmentation is the cornerstone of customer insight and understanding across the modern marketing organization.

Segmentation can be used to build and manage stronger, longer, and more profitable customer lifetime relationships. Different customers are motivated by varying attributes, treatments, and interactions. Segmentation categorizes or classifies items or subjects into identifiable groups that share similar features. Features can be based on demographics, purchase behaviors, clickstream patterns, channel interactions, and the list goes on. If a company does not partition their customers into segments, they might be likely to treat all customers the same.

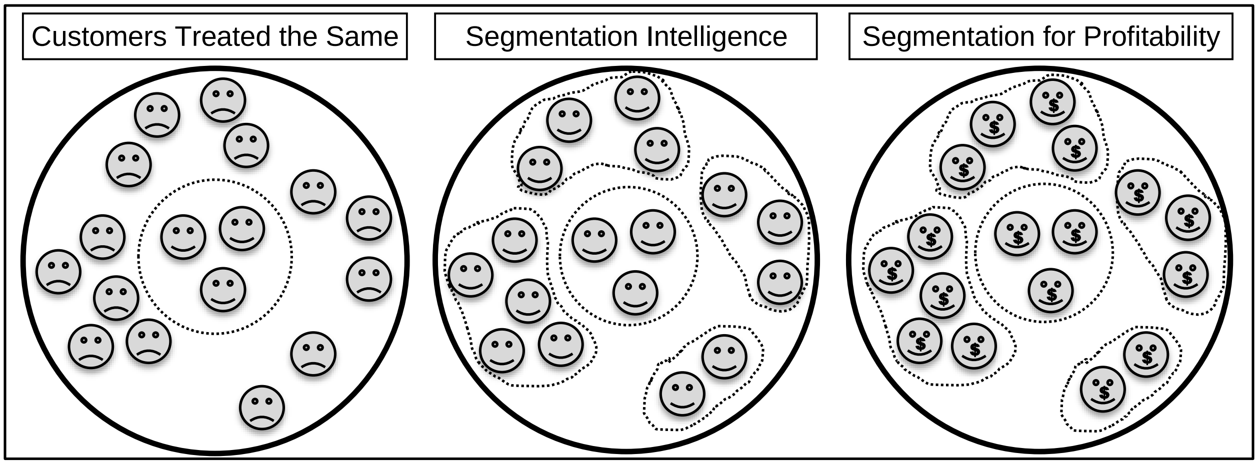

Figure 2: Customer Segmentation Intelligence (CSI) For Increasing Profits

We can use a visual image to demonstrate differences among customers. Notice from the left group in *Figure 2* that only the average customers inside the center dotted circle are happy. The company is treating all customers the same as if they were all average. However, that applies to only that small set in the center dotted region. The others are sad because their needs and wants are not being addressed. If you intelligently segment your customers into groups that can be treated differently and in accordance with their needs and wants, you should be able to make almost all customers happy as shown in the middle group. And your profits should grow as your new and existing customers become more loyal advocates for your brand, as in the group on the right.

Segmentation is both an art and a science. While creative visionaries and data scientists are both tremendous organizational assets within a team, it is the alliance between these two segments that will push marketing forward. In this paper, we are providing the perspective and techniques around the science of segmentation. Understanding your business, your customers, and how you tell a story with data provides the art. Influencing decision-makers within an organization isn't easy, and if they do not understand the analysis, nothing will ever change. There are people who are good at creative marketing strategy, and there are people who are good at marketing analytics. However, there aren't many people who can toggle between the two and serve as the translator who inspires both sides.

Part of the challenge in modern marketing is the continued struggle with more complex and richer data streams. Although issues with data management and interpretable analytics do exist, data-driven marketers have an opportunity to establish segmentation as the golden mean for customer understanding—resting between the two extremes of treating all customers the same and one-to-one personalization.

WHAT IS THE BEST SEGMENTATION?

There are many different approaches to segmentation, some are simple, and others are more complex. However, they all have one thing in common: Segmentation is used to differentiate customers into homogeneous groups so that they can be targeted with personalized treatments. By receiving relevant treatments, customers are more likely to behave in a profitable manner as well as being more loyal, subsequently driving customer lifetime value in a positive direction.

There are two major ways to increase customer profit. The first is by increasing market share—you get more customers. The second is by increasing share of wallet—you get more money from each customer. In both cases, segmentation enables you to treat customers differently based on the segment that each individual belongs to.

So, is there a “right” number of segments? We segment in order to manage our marketing initiatives spanning millions of consumers. Our intent within the discipline of Customer Relationship Management (CRM) is to nurture relationships with each customer, optimizing the individual’s experience with the brand. It is extremely challenging to conduct one million discrete campaign objectives. So we segment as a compromise between marketing to each customer individually and treating all customers the same.

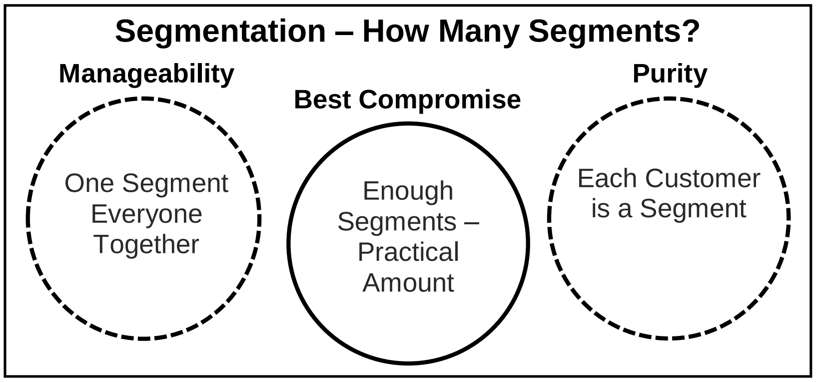

Figure 3: How Many Segments?

As shown in *Figure 3*, you want to have enough segments to treat customers according to their needs and wants without the effort becoming unmanageable. Furthermore, you might decide to have multiple segmentation models based on different dimensions:

- Segmentation strategy based on purchase history
- Segmentation strategy based on digital interactions
- Segmentation strategy based on demographics

Of course, it is possible and worthwhile to create a unified segmentation strategy across all dimensions, sometimes referred to as ensemble segmentation.

Methodologies can be executed using business rules (nonquantitative), analytical methods (quantitative), or some combination of the two. In this paper, we will define the term segmentation as a process of grouping observations or customers. Thus, analytical methods that encompass both supervised (classification) and unsupervised (clustering) approaches will be defined as a subset of segmentation that uses a statistical approach to analyze the data relationships.

HOW SHOULD YOU USE SEGMENTATION?

There are two main business reasons for segmenting your customers:

1. Treat and manage customers in each segment differently
2. Use segments in further analysis

By segmenting customers (or prospects), the expectation is that different groups will respond uniquely to targeted communications, incentives, or offers. Ideally, marketers will want to interact with each segment in a manner that best matches the profile characteristics of that audience. Because different attributes can help explain why customer groups behave differently, one can also run predictive analytic models within each segment. For example, you can predict the likelihood of responding to an offer to buy a product. However, since you might want to sell products to different segments, you would be better off building propensity models for each segment. This usually increases the signal-to-noise ratio, or predictive accuracy, of each model compared to building one model for all customers.

Ultimately, segmentation, like any analytical exercise within an organization, needs to prove its value. To save your segmentation projects from failing, you must be clear on how your organization will act differently with the results. Early in your segmentation journey, you should do the following:

Identify relevant use cases. To succeed with a segmentation effort, you need to determine the right use cases. "Understand our customers better" or "improve prospect targeting" are not specific enough goals to guide data scientists or marketing analysts. Push yourself to exploit specified topics, such as "improve email (or direct mail) channel performance through analytically derived segments," or "orchestrate highly relevant website or mobile app experiences through segmentation insights derived from actionable clickstream behaviors."

Explore your internal and external data inputs. With your use cases selected, you can prioritize investments in collecting and stitching data together to support segmentation. Leveraging different data sources is critical for getting a comprehensive picture of your customers, so strive for as many of the following as possible:

- Customer relationship management data (CRM)
- Point-of-sale data (POS)
- Media channel campaign data (if it isn't available in-house, extracts are typically available from your ad agency or attribution vendor partners)
- Web visitor behavioral data
- Mobile app visitor behavioral data
- Third-party data appends (both traditional offline sources as well as extracts from your marketing service provider or digital data management platform partners)
- And the list goes on…

As you can see, we are living in a world of ever-expanding data sources, and understanding your options is immensely important. Even if you don't have access to some of the sources listed above, it's better to work with what you do have access to, prove value through actionable segmentation, and make investments into additional data over time.

Target actionable projects. There isn't much more to say here. If you can't take action on the insights you share with marketing, then what was the point of the exercise? Activate segmentation by identifying multiple projects to apply segment knowledge. Senior leadership wants results that matter, and this is an extremely important consideration in your efforts to succeed. Remember at the end of the day, any form of marketing or advertising is a bet, and will result in positive or negative revenue.

EXAMPLES OF SEGMENTATION IN INDUSTRY

Consider the use of segmentation in a direct marketing scenario where you have ten million customers in your marketable universe. The brand's product offering is so large and varied that it cannot offer all of it in a single creative package, such as a catalog. Also, treating all customers the same by offering only one creative package is ineffective because the catalog would have to appeal to all segments, and each customer would have to search the huge catalog to find the desired items. Therefore, the brand uses their customer relationship management data by segmenting the marketable audience into groups based on their demographic and purchase patterns. Each customer segment then is provided with a catalog designed specifically for them. A catalog for a teen segment would be very different from the one designed for middle-aged adults. This approach is manageable and has a higher propensity to increase customer responsiveness to the brand's offerings.

Digital marketing leverages segmentation to provide more individualized recommendations similar to what direct marketers have done in the past. Here are a few modern examples across the stages of the customer lifecycle:

- Acquisition: Digital marketers target prospects to add to their brand's sales funnel by analyzing different segments of current customers, and then by using those profiles to influence how they target prospective "look-a-likes" through retargeting, display, and social media campaigns. This is based on the well-known principle that "birds of a feather flock together," and proves to be an effective method across any industry when the majority of visitors to your website or mobile app are unidentifiable.
- Upsell/cross-sell: If I am a fan of a sports team, and I have purchased tickets in the past for gaming events, the team's organization has an interest in me returning to future games. Segmentation can be used for setting up next-best-offer campaigns, which can focus on a variety of use cases. For example, if I belong to a family segment, perhaps the sports team can entice me to return with email campaign offers to meet the team mascot at the next home game (the kids would love that), as opposed to offering me discounts for alcohol purchases, which might not resonate with me, but would with a young, singles segment.

- Retention: Hospitality brands are obsessed with maintaining affinity with their guests, as many options exist today on where to book a hotel room. Different segments of behavior indicate why one guest prefers a particular brand, from pricing, environment, bed quality, workout facilities, and spas. How do you keep segments of guests coming back? Personalization is huge, and the more you understand about segment preferences, opportunities to drive relevance on future website, mobile app, and on-property experiences can make a significant difference.

Maintaining positive brand sentiment is another aspect where segmentation can benefit your organization's initiatives. For example, in the health care industry, it is important to understand the needs and wants of different groups to provide the desired benefits at the right cost. Some people want more preventive health care because they take better care of themselves. Other people want protection against catastrophic events, but are less concerned about regular health care. Some are willing to pay for everything. If the company segments the customers into meaningful groups, the results will be health care that better meets the customer desires and higher profitability for the providers and administrators.

GENERAL APPROACHES FOR SEGMENTATION

There are four general methods to group or segment customers:

1. Business rules – This technique centers on a qualitative, or non-quantitative, approach leveraging various customer attributes that are conceptualized through conversations with business stakeholders and customer focus groups to gather pointed data. This information represents consumer experiential behavior, and analysts will assign subjective segments for targeted campaign treatments directed by specific, predetermined values of the attributes used.
2. Quantiles – This approach creates segments based on cut points that divide the marketable population into equal size groups (or bins), typically derived from assessing the univariate statistical distributions of the available customer attributes. One can use quantiles from various attributes such as Recency, Frequency, and Monetary (RFM) of customer behavior to create a more in-depth and meaningful segmentation.
3. Supervised classification – Referred to as a family of algorithmic pattern analysis approaches, supervised segmentation delivers homogeneous segments that can be profiled, and informs targeting strategies across the customer lifecycle. The use of the term "supervised" refers to specific data mining (or data science) techniques, such as decision trees, regression, gradient-boosting, or neural networks. From a marketing perspective, decision trees are the most popular technique due to their ease of interpretation. The key differentiator in supervised segmentation is that the analysis requires a dependent (or target) variable, usually a 1-0 (or yes/no) flag-type variable that matches the objective of the segmentation, and profiles different audiences with varying levels of propensity aligned to that objective (offer response, cross-sell, churn, and so on).
4. Unsupervised clustering: Unsupervised algorithmic segmentation, such as k-means or hierarchical clustering, association/apriori, principal components or factor analysis, point to a subset of multivariate segmentation techniques that group consumers based on similar characteristics. The goal is to explore the data to find intrinsic structures. K-means cluster analysis is a very popular technique for interdependent segmentation, in which all applicable data attributes are simultaneously considered, and there is no splitting of dependent (or target) and independent (or predictor) variables (that is, there isn't a 1-0 or yes/no flag type variable to bias the formation of the segments).

Which approach works the best? That depends on two factors: the business goal and customer attributes. The business goal must drive which segmentation process will be used. However, that will depend on the customer attribute data that is available as well as the importance of the customer attributes in building the segmentation models.

In conjunction with these methods, SAS Visual Analytics works well to manage and visualize business rule and quantile segmentation. SAS Visual Statistics works nicely to analyze and visualize supervised and unsupervised segmentation through explanatory modeling and clustering.

In the next sections, you will read a synopsis of each approach with examples, advantages versus disadvantages, and technology applications with SAS Visual Analytics and SAS Visual Statistics. This should enable you to make a more informed decision about which segmentation method will work best for you.

BUSINESS RULES

The simplest and most direct method of segmentation is the application of business rules. When we use business rules, we assign each customer to a group according to predetermined classes. Examples of applying business rules might include grouping based on demographics of age, gender, income, education attained, and so on. Grouping can also be done based on the interaction that a customer has with the company, such as types of merchandise or service purchased, amount of money spent, when, how, and where the purchase is made (online, brick-and-mortar store, phone), and more. Sometimes customer purchase behavior known as RFM segmentations can be a foundation method used to group customers. This can be done by applying business rules for creating the groups. For example, customers having shopped within the past two months, twenty times over the past year, and spent an average of $50 or more per week are in a “Gold Segment”. You also could use univariate statistics to create quantiles, which will be covered in the next section.

Prior to point-and-click visualization through SAS Visual Analytics and SAS Visual Statistics, you would often look at segmentation results through tables of numbers comparing the various attributes across segments. Now, with visualization, you can quickly and easily view with heat maps the differences among segments with any attribute that is of interest.

The following example using SAS Visual Analytics demonstrates the visualization of business rule segmentation.

On the left, we choose the variables that we want to use for the segmentation—age_grp (age group) and F or M (gender). Then we choose the measure (1 if responded to test mailing) to see how each segment has performed. The segments of respondents to a test campaign varied from an average of 5% to 15% response rate. Next, we choose the heat map visualization because it most easily displays the differences among the segments.

By these few clicks in the SAS Visual Analytics interface, you can now see the heat map of customer segmentation in *Figure 5*. Note that you can quickly see the age and gender groups that are more likely to respond to a mailing versus those that are less likely. Further, the young respondents might have a different reason for responding than the 45-48 year olds.

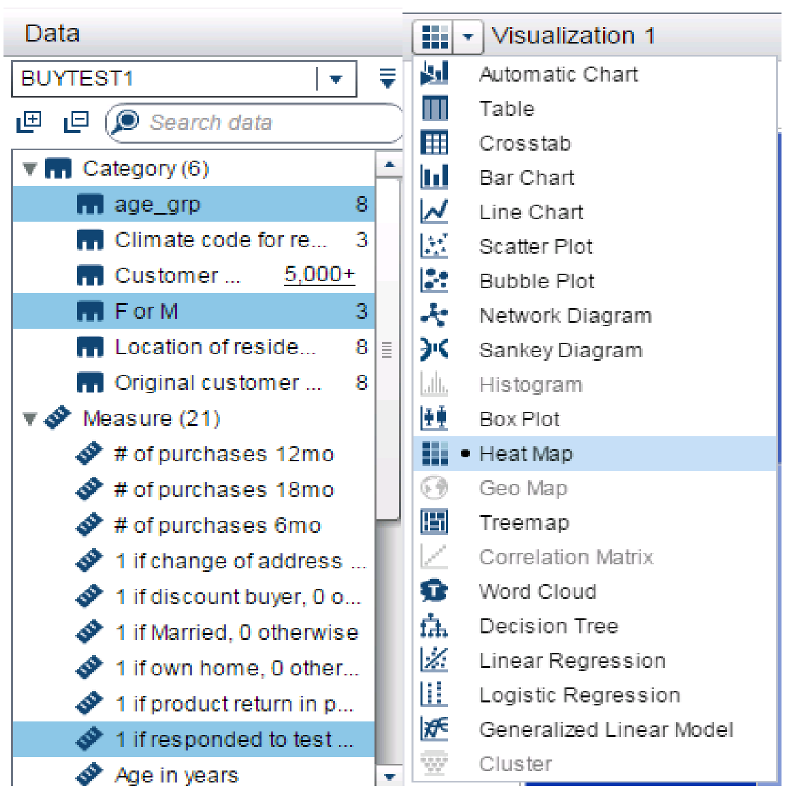

Figure 4: SAS Visual Analytics Parameter Selection Windows

By creating segments this way, you can communicate differently with each group.

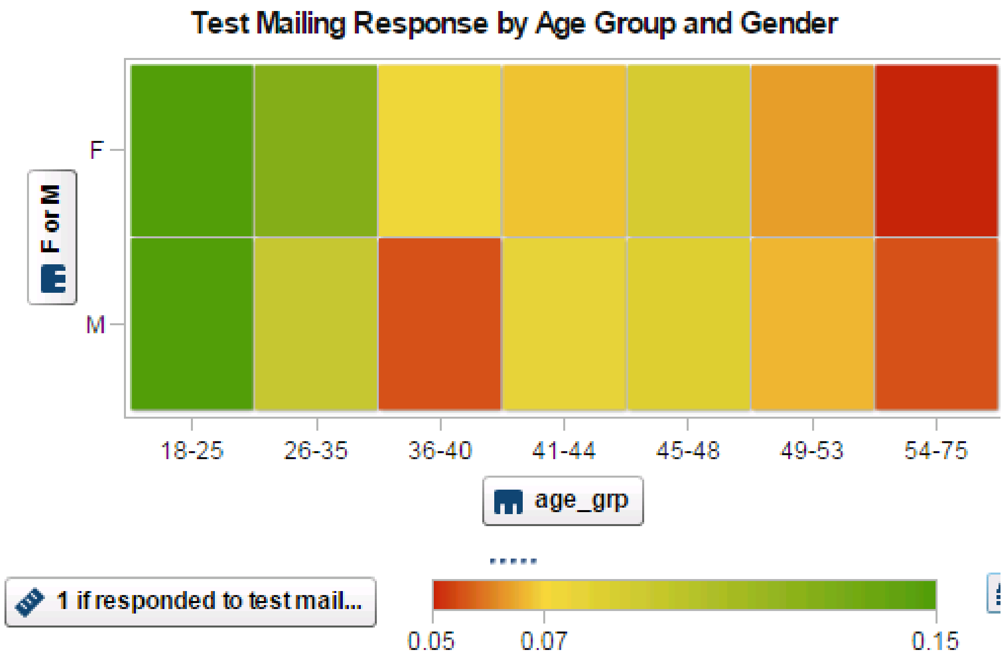

Figure 5: Heat Map For Business Rule Segmentation In SAS Visual Analytics

However, the assumption is made that the differences among the groups are meaningful. This implies that the customers within each group will behave differently based on their group membership resulting in increased loyalty as well as increased customer acquisitions.
There are advantages and disadvantages to using business rule segmentation.

Business Rules Advantages:

- Consistent with business goals
- Easy to apply

Business Rules Disadvantages:

- Might not reflect the reality of customer behavior
- Some segments might not be significantly different from one another

QUANTILE MEMBERSHIP

Creating segments based on quantile membership uses attribute frequencies to determine segment membership. With business rules, the grouping definitions are predetermined. With the quantile approach, univariate data analysis determines the groupings. The quantile approach is often used to create groups using customer purchase behavior attributes such as Recency (that is, when did customers last shop?), Frequency (that is, how often did customer shop?), and Monetary (that is, how much did customers spend?). Collectively these are known as RFM segmentations. Monetary value can be just the amount of money spent per purchase or in some cases per customer within a time period or even over the total customer tenure. However, the latter example of calculating monetary segmentation might conflict with frequency. The time period chosen is very important. You want to have a long enough time period to include all “active” customers. Yet, you want a short enough time period that the historical behavior provides a reasonable pattern going forward. Calculating RFM can be done by using univariate statistics to create quantiles, which will enable grouping (that is, top 10% of customers on the basis of RFM are the “Gold Segment”).

Using the RFM attributes to create segments, you can divide each metric into a number of quantiles. Experience has shown that dividing the metrics into five quantiles (also known as quintiles) provides the best amplification of signal without introducing too much noise into the segments. You could use just three groups for each metric. However, that might not provide sufficient discrimination among customers. Using seven groups provides little additional understanding of the customer population while adding to what must be interpreted. Therefore, five quantiles is recommended.

For example, suppose you had the customer data for a home improvement company. You examine the purchase dates for the last year. You also have total purchase amounts for the last year. Because you have products with a wide range of prices, you might be more interested in the recency and monetary (total purchases over the past year) aspects of your customers rather than the frequency. For each customer, you have their sum of all purchases and last purchase date. You can quickly transform each variable into one of five ranked groups. Crossing the five monetary groups with the five recency groups provides us with twenty five groups. With SAS Visual Analytics, we can visualize the quantiles of recency and total purchases using the heat map again to ascertain the “best” and “worst” customers.

This heat map from SAS Visual Analytics looks at RFM with only recency and monetary. It illustrates how from a marketing perspective we might group segments created with quantiles. Here we have 31,000 customers in one region. We have divided the recency into quantiles based on the number of days since the last purchase. We notice that 20% of our customers have made purchases in the last week.

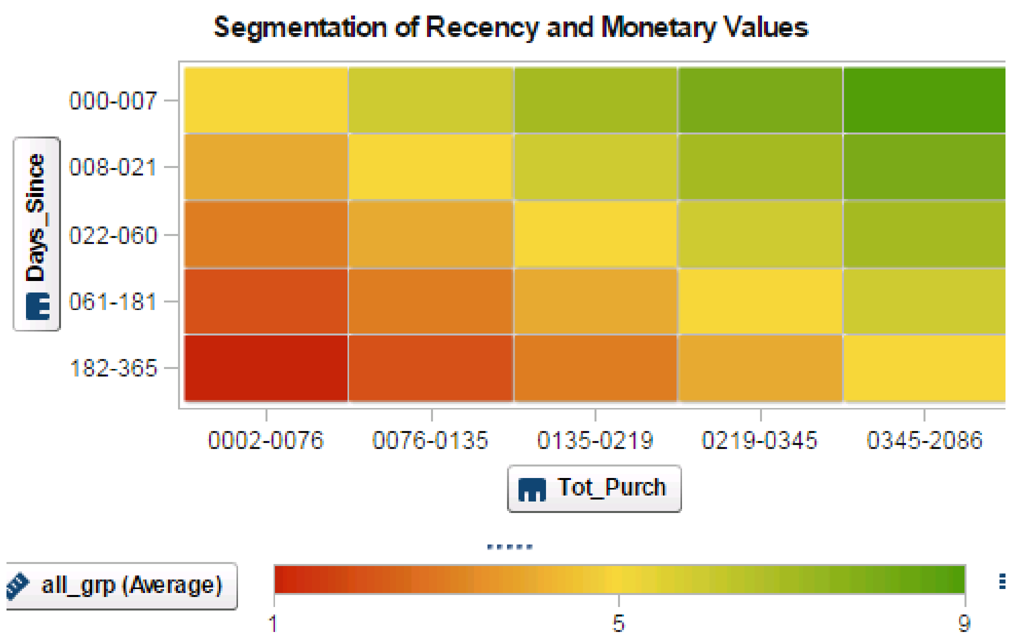

Figure 6: RFM Heat Map For Quantile Segmentation In SAS Visual Analytics

Another 20% purchased between two and three weeks ago, and so on. For monetary, we notice that 20% of our customers have made total purchases of over $345 in the last year while the lowest 20% purchased less than $76 in the last year, and so on. We might choose to treat the twenty five segments as nine groups from the least valuable colored red to the most valuable colored green. Those customers who have not shopped in the last six months and spent a total of less than $76 are categorized as the least valuable while those who shopped within the last seven days and have spent a total of more than $345 are categorized as the most valuable. Average customers would be those who have spent less than $76 even when shopping in the last week which is the same as customers who have not shopped in the last six months, but have spent over $345. Obviously, the value placed on any segment with RFM is related to the business value a company places on these various groups.

The use of RFM in marketing is to keep the strong bond with good customers and to improve the bond with customers who have weaknesses in one or more areas. However, if you think about the life cycle of a customer, you might want to keep those customers who have mediocre RFM values because in the long term they might prove to be more valuable. Therefore, RFM does not answer marketing issues. It simply provides the segmented data with which to make better marketing decisions. RFM is not predictive; it is only descriptive. Of course, it assumes that past behavior is indicative of future behavior. However, if you change the interaction with the customer, the future behavior might change.

Quantile Membership Advantages:

- Clearly reflects segments based on chosen metrics.
- Easy to apply.

Quantile Membership Disadvantages:

- Does not provide a mechanism to forecast behavior as a predictive model might.
- Metrics for segmentation pre-chosen. Better segments might be obtained with a different set of attributes.

SUPERVISED CLASSIFICATION

The usage of supervised segmentation requires an analyst to identify a dependent (or target) variable that matches the objective of the exercise, and to explore all available attributes (or independent variables) to identify which are important in describing unique profiles of different audiences within a marketable population. Critics of this approach argue that the resulting analysis is actually a predictive model rather than a segmentation model because of the probability prediction output. The distinction lies in the use of the model. Supervised segmentation is classifying customer bases into distinct groups based on multidimensional data and is used to suggest an actionable roadmap to design relevant marketing, product, and customer service strategies to drive desired business outcomes. As long as we stay focused on this premise, there is nothing to debate.

As previously mentioned, within the family of supervised segmentation techniques, decision tree analysis is heavily leveraged in marketing analytics. Decision trees are relatively easy-to-understand and explain, particularly to a non-technical audience, making the insights actionable. Decision trees fall into two general categories:

1. Classification trees
2. Regression trees

If the business objective is categorical (that is, nominal or ordinal), then a classification tree is used to predict the probability of a particular outcome based on a set of predictors (or independent variables). If the business objective variable is continuous (or numeric), then a regression tree is used to predict the mean (or average) of the dependent (or target) variable.

Examples of how marketing analysts leverage decision trees in everyday life include the following:

1. What type of audience segments should we advertise to for a new movie?
2. What geographic segments will vote for a particular candidate in an election?
3. What type of buyer profiles are best to target for a new housing and real estate development?
4. What types of travelers have an affinity to a particular airline?
5. What is the risk profile of an applicant that has a high propensity to default on a loan?

Each of these examples presumes a marketing objective whether that be an offer, a desired behavior, or a profile connected to a likelihood to buy, vote, default, etc. Decision tree analysis consists of a set of conditional rules, based on decision thresholds. The tree is essentially a series of nested if-then statements that lead to a classification (or segment).

Here are a couple of credit risk assessment examples:

- If income is > 100,000 and credit rating is > 600, then credit risk is low; but if credit rating is < 600, then credit risk is high, and…
- If income is <=100,000 and credit rating is < 700, then credit risk is high; but if credit rating is > 700, then credit risk is low.

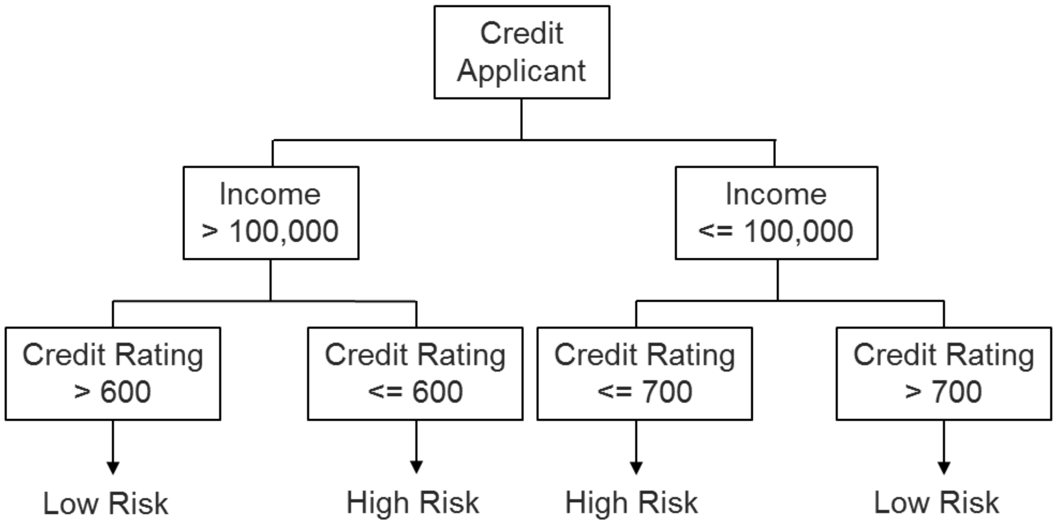

Figure 7: Simple Illustration - Decision Tree Analysis

The objectives of decision tree analysis are as follows:

- Segmenting the marketable population into groups that are as homogeneous as possible and maximizing the difference in the customer behavior of those groups, with respect to the business objective.
- As the decision tree grows, successive splits of the marketable audience produces a branched structure of rules and groups, which is the model of the data.
- The rules are readily expressed in English making them easily understood, and can also be expressed in SQL (to retrieve or score matching records). This is what makes "look-a-like modeling" possible!
- As supervised models, decision trees can be used for classification**,** estimation, and prediction. With respect to segmentation, classification is the application to focus on.
- Different variants of the decision tree algorithm exist, such as CART, CHAID, C5.0, and C4.5. Although differences exist within the math of these approaches, all types are respectable options for supervised segmentation use cases.

And here is a deeper dive into the objectives of decision tree analysis:

- Decision trees are a machine-learning method that can split customer data into smaller and smaller segments, which are increasingly "pure" (or unique).
- The algorithm treats each identified (or partitioned) audience segment independently. To find a new segment (or split), the algorithm tests potential splits based on all available customer data, and prioritizes the newly identified segment based on a statistical assessment, typically the LogWorth or G^2 test.

- In doing so, the decision tree chooses the most important variables for the supervised segmentation task, and identifies the audience segments based on specific values of those variables that correlate to the business objective.

Let's transition now and explore a real-life use case for modern marketing leveraging SAS Visual Statistics within the context of digital segmentation and predictive marketing.

- Business Problem: What drives sales conversion behavior from visitors to a brand's website or mobile app?
    - Decision tree analysis can be used to provide a supervised modeling approach to data-driven segmentation, in correspondence to understanding drivers of business conversions on a website or mobile app. The insights can then be used to influence downstream marketing personalization and targeting campaigns.
    - After applying this technique, marketing analysts can deliver a visual representation of the segments to help explain the results to nontechnical stakeholders.

For this use case, we have recorded a video demonstration leveraging SAS Visual Statistics to showcase the analysis. Please access the recording here:

https://www.youtube.com/watch?v=J_EtjCWcI84&index=6&list=PLVBcK_IpFVi86GuTpnzE4I1lmYOYP_f6e

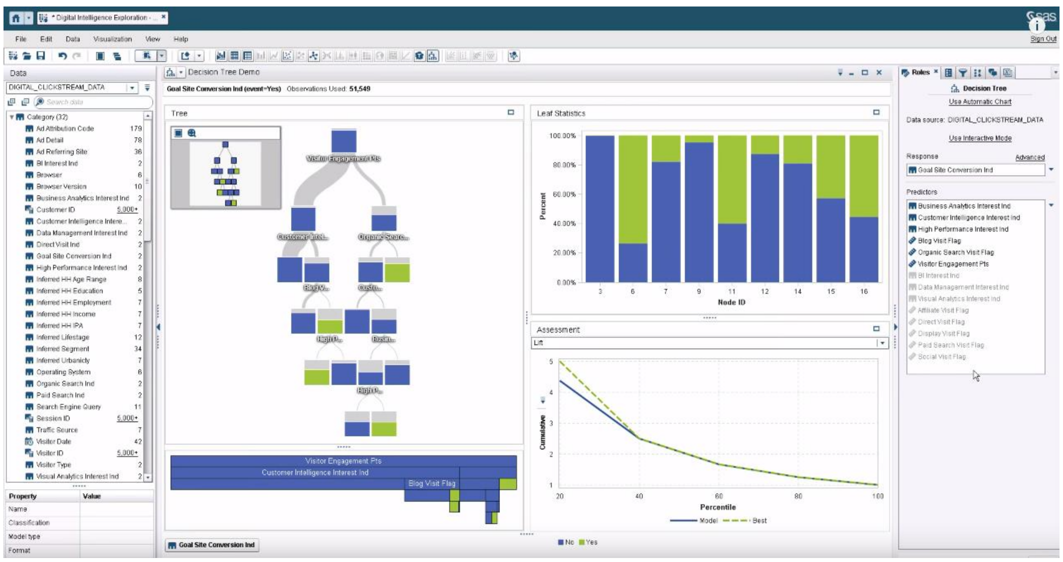

Figure 8: Decision Tree Analysis In SAS Visual Statistics

Advantages of supervised segmentation include the following:

- Mapping segments to a specific business objective
- Algorithmically derived while being easy-to-explain to a non-technical audience

Disadvantages of supervised segmentation include the following:

- As with all analytical exercises, supervised segmentation requires additional effort to oversee and ensure that modeling insights remain accurate over time.
- Model cross-validation and in-market testing are considered best practices.

UNSUPERVISED CLUSTERING

Recall that supervised segmentation is based on finding groups of customers that differ with respect to some target characteristic of interest. In other applications, we may want to find groups of customers that are not driven by pre-specified target characteristics. Do our customers naturally fall into different groups? This information can be useful for many reasons.

For example, we might want to step back and consider our marketing efforts more broadly. Do we understand who our customers really are? This idea of finding natural groupings in the data is what we call unsupervised segmentation.

The basic idea is that we want to find different groups of consumers, where the consumers within a specific group are similar, but the consumers between groups are not so similar. Since the usage of unsupervised segmentation does not require an analyst to identify a dependent (or target) variable that matches the objective of the exercise, it is fair to make the interpretation that the results are more data driven, hence more natural and better suited to the underlying structure of the data. This advantage is also its major drawback: it can be difficult to judge the quality of clustering results in a conclusive way without running live campaigns. In marketing, A/B testing can be useful in addressing this.

As previously mentioned, within the family of unsupervised segmentation techniques, k-means clustering analysis is a very common application in marketing analytics. The main idea behind this technique deserves some description.

In k-means, the “means” are the centroids, represented by the arithmetic means (or averages) of the values along each dimension (or attribute) for the members in the cluster (or segment). So in the figure below, to compute the centroid for each cluster (or segment), we would average all the *x* values of the points in the cluster to form the *x* coordinate of the centroid, and average all the *y* values to form the centroid’s *y* coordinate. Generally, the centroid is the average of the values for each attribute of each member (or customer) in the cluster.

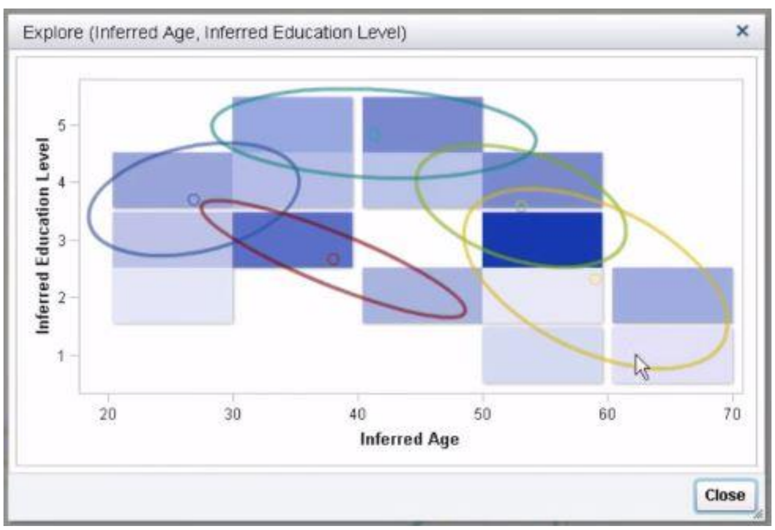

Figure 9: Cluster Matrix Visualization In SAS Visual Statistics

The *k* in k-means is simply the number of segments that you would like to find in the data. The beginning of a k-means analysis starts with a desired (or subjective) number of clusters *k*. So, in the figure referenced above, the analyst would have specified k=5, and the k-means clustering method would return the following:

- The five cluster centroids when the clustering algorithm had completed iterating.
- Information on which of the data points (or customers) belong to each cluster.

The k-means algorithm for finding segments is simple and effective, and therefore is worth highlighting:

1. The algorithm starts by creating *k* initial cluster centers, usually randomly.

2. Next, as more data is analyzed, the segments corresponding to these cluster centers are formed, by determining which is the closest center to each point or observation.
3. For each of these clusters, its center is recalculated iteratively as more data is assessed by finding the actual centroid of the points in the cluster.
4. The cluster centers typically shift while the process continues. Since cluster centers will shift, the algorithm continually recalculates which points belong to each cluster. After reassignments, the cluster centers might shift again and again.
5. The k-means procedure keeps iterating until there is no change in the clusters, and the process terminates.

Keep in mind, there is no guarantee that a single run of the k-means algorithm will result in a good segmentation analysis. Although a single clustering run will find a local optimum—a locally best clustering—this is dependent upon the starting centroid locations. For this reason, k-means is usually run many times by a marketing analyst, starting with different random centroids each time. The results can be compared by analyst examination, or by a numeric measure such as the clusters' distortion, which is the sum of the squared differences between each data point and its corresponding centroid. In the latter case, the clustering with the lowest distortion value can be deemed the best segmentation.

One best practice when executing a k-means clustering suggests sorting the population to be clustered segmented in descending order by the squared distance from the global mean over all the attributes used. With this approach, the first cluster begins with the observation or customer most distant from the population mean. Each additional customer added is the next farthest from the population mean, and the process proceeds. This approach often results in clusters (or segments) being more homogenous (or pure).

A common concern with k-means clustering is how to determine a good value for *k*. One answer is simply to try different *k* values and see which iteration produces quality results. The value for *k* can be decreased if some clusters are too small and overly specific, and increased if some clusters are too broad and diffuse. Wikipedia's article "Determining the number of clusters in a data set" describes various metrics for evaluating sets of candidate clusters.

With that stated, let's dive into a marketing use case leveraging SAS Visual Statistics within the context of unsupervised segmentation and k-means clustering:

- Business Problem: Can we develop better products, better marketing campaigns, better sales methods, or better customer service by understanding the natural subgroups of our customer data?
- K-means clustering analysis can be used to provide an unsupervised modeling approach to data-driven segmentation. In this use case, we will aim to improve our understanding of inferred customer demographic profiles (typically gathered from a digital data management platform or marketing services provider partner) from your brand's website or mobile app visitors. The insights can then be used to influence downstream marketing personalization and targeting campaigns, as well as leveraged in two-step modeling processes where the identified segments can subsequently be used in a group-by regression or decision tree propensity model.
    - After applying this technique, marketing analysts can deliver nontechnical stories of the identified segments to help explain the results to business stakeholders.

For this use case, we have recorded a video demonstration leveraging SAS Visual Statistics to showcase the analysis. Please access the recording here:

https://www.youtube.com/watch?v=VBxdoL84xrM&list=PLVBcK_IpFVi86GuTpnzE4I1lmYOYP_f6e&index=7

Figure 10: Clustering Analysis In SAS Visual Statistics

As you can see in the video (and in real-world analysis), some clusters are interesting and thematically consistent, while others might not be. There is an old viewpoint in statistics: Correlation is not causation, meaning that just because two things co-occur doesn't mean that one causes another. We shouldn't expect every cluster to be meaningful and interesting, but the discovery of high-quality segments creates data-driven evidence to take action.

Also, in the last minute of the video, we highlighted the ability to derive a cluster ID variable, which allows an analyst to enrich customer data by tagging individuals with segment assignments. Why is this useful? Because now the analyst can use a second-step supervised modeling approach, such as a logistic regression or decision tree, and build multiple predictive models by segment, targeting a specific business objective. Instead of building one model, an analyst can build five, ten, or twenty supervised models in a single-click, and typically derive more accurate (and actionable) results.

We often observe that what drives the customers in one cluster to respond to a communication or offer is often different from those in another cluster. A cluster of young, low-income people might be price sensitive, while an older more salaried group might be interested in higher-quality products.

Here is a recorded video demonstration of SAS Visual Statistics in the context of logistic regression, leveraging the group-by capability for different visitor segments of a brand's website:

https://www.youtube.com/watch?v=Z8S9bsx11e0

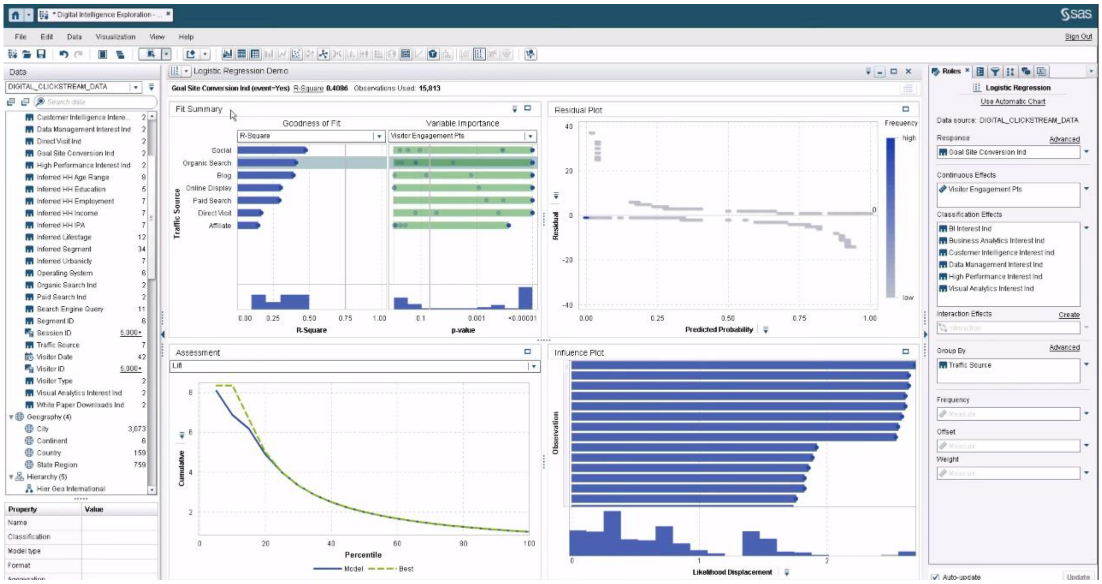

Figure 11: Logistic Regression Analysis In SAS Visual Statistics

Seven propensity models created and run in a single click. That's the power of SAS Visual Statistics, and now we can leverage these results across the customer lifecycle. In the demonstration, we focused on acquisition, and we now have unique insights to target "look-a-likes" across the different segments (which can be derived from unsupervised segmentation, or even non-quantitative methods like business rules).

Nevertheless, unsupervised clustering is often a useful tool to uncover structure in our data that we did not foresee. Clusters can suggest new and interesting marketing campaigns, as well as other data mining opportunities.

Advantages of unsupervised segmentation include the following:

- The results are more data driven, and group customers into organic, or more naturally occurring, behavioral groups.
- The results can be coupled into a two-step analytical modeling process, taking advantage of the benefits from both unsupervised and supervised algorithms.
- The results are algorithmically derived while being easy-to-explain to a non-technical audience.

Disadvantages of unsupervised segmentation:

- There is no guarantee that a single run of the k-means algorithm will result in a good segmentation analysis. It typically requires multiple passes by a marketing analyst.
- When using K-Means Clustering, it can be moderately difficult to determine a good value for *k* (or number of segments).
- As with all analytical exercises, requires additional effort regarding to oversee and ensure that modeling insights remain accurate over time.
- In-market testing, such as A/B testing, are considered best practices in comparing the performance of algorithmically derived segments versus control groups.

CONCLUSION

Segmentation enables you to treat your customers relevantly according to their characteristics. In this way, you can improve your ability to meet your customers' needs and wants. So, which segmentation should you use? Determine what you want to accomplish and use that to decide on the method. If you have certain business rules that must be followed, the business rule segmentation can be used. If you want to follow an RFM approach or some other metrics that lend themselves to the use of quantiles, then that is a suitable option. If you have a business objective, such as sales conversions or ad-clicks, supervised segmentation is highly applicable. The last approach, unsupervised segmentation with k-means clustering, is easily the most flexible and powerful approach and requires the fewest assumptions about your customer population.

SAS Visual Analytics and SAS Visual Statistics represent a new category of interactive and collaborative technology to provide a path to be curious and innovative—both for technical and non-technical stakeholders. Marketers are imaginative, and are constantly pushing to analyze new and exciting data sources (that is, clickstream, social, IoT wearables, and so on), which require the ability to scale to very large amounts of information. However, what is compelling is the ability of SAS Visual Analytics and SAS Visual Statistics to perform sophisticated analysis, and produce visualizations to support data-driven storytelling to ensure that brands understand and take action on insights.

With the availability of visualization and running your analyses in an interactive, point-and-click interface, you can more easily use the more sophisticated segmentation approaches of supervised classification and unsupervised clustering. You can always add in your business rules and quantile approaches if you find them to be useful. However, you should not feel restricted to those because you don't have the statistical capabilities to process your data. SAS Visual Analytics and SAS Visual Statistics offer you the technology to create market-driven segments, enabling a more powerful marketing strategy that is both comprehensive and nimble. As you incorporate algorithmically-derived segmentation, and in some cases, predictive scoring by segment into your marketing strategies, you can quickly take results from one campaign and influence future campaigns with increased intelligence and personalization.

You can try more than one method to assess the "goodness" of different approaches in your environment. You can even combine different methods, such as ensemble models, if the situation warrants. The authors strongly encourage the use of segmentation prior to building predictive models. You might want to compare results from building predictive models within clusters to the predictive model over the whole population. You might also want to use the cluster segments as an additional input variable (or predictor) to your predictive model. No matter which segmentation method is used, by applying Customer Segmentation Intelligence (CSI) and creating and visualizing your results within SAS Visual Analytics and SAS Visual Statistics , you will be able to simultaneously increase relevance with your customer's needs, and improve profitability.

REFERENCES

1. Baer, Darius S. 2012. "CSI: Customer Segmentation Intelligence for Increasing Profits." *Proceedings of the SAS Global Forum 2012 Conference*. Cary, NC: SAS Institute Inc. Available at http://support.sas.com/resources/papers/proceedings12/103-2012.pdf.
2. "Forrester Wave™: Customer Analytics Solutions, Q1 2016." Forrester Research, Inc. March 2016.
3. Collica, Randall S. 2011. *Customer Segmentation and Clustering using SAS® Enterprise Miner™, Second Edition*. Cary, NC: SAS Institute, Inc.

RECOMMENDED READING

Customer Segmentation and Clustering using SAS® Enterprise Miner™ Second Edition

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Name: Darius Baer
Enterprise: SAS Institute, Inc.
Address: 420 Edison Place
City, State ZIP: Superior, CO 80027
Work Phone: 919-531-9510
Cell Phone: 303-204-7015
E-mail: darius.baer@sas.com
LinkedIn: https://www.linkedin.com/in/darius-baer-614a8b
Web: www.sas.com

Name: Suneel Grover
Title: Principal Solutions Architect / Professorial Lecturer
Enterprise: SAS Institute, Inc. / The George Washington University
City, State: Washington D.C.
E-mail: suneel.grover@sas.com
LinkedIn: https://www.linkedin.com/in/suneelgrover
Twitter: @suneelgrover
Web: www.sas.com