

Retail Clustering Methods: Achieving Success with Assortment Planning



Executive Summary

Retail assortment planning is a topic that has increasingly gained attention and momentum within retail executive suites and merchandising solution vendors' future development and enhancement plans. With the wider interest and use of “big data” and analytics – as well as more robust tools for supporting these capabilities and increasingly demanding and fickle customers – wider attention to assortment planning is inevitable. A foundational element of effective assortment planning is the ability to appropriately cluster stores and channels to maximize sell-through and margin potential. However, this key capability is rarely given top priority – often viewed as a mundane, analytical effort and is assumed to be “built in” to the assortment planning solution.

Effective clustering provides the ability to unleash the true potential of assortment planning capabilities, bringing with it significant financial benefits in terms of sales, margin, and inventory utilization – as well as improved customer satisfaction, due to being better able to provide the “right” mix of products for customers, across locations and channels.

In this Point of View, The Parker Avery Group discusses how clustering for assortment planning is an intricate undertaking with a variety of approaches and elements to consider. Granted, there are simple, straightforward clustering methods, but these tend to have significant shortcomings, and typically fail to create assortments that drive meaningful results. Conversely, more sophisticated approaches usually require more skilled resources, solid data integrity, and appropriate supporting solutions to take advantage of the potential these methods can deliver. We will explore ten different clustering approaches in depth and highlight the advantages, disadvantages and under which circumstances each should be used. This understanding, coupled with clearly defined assortment planning objectives, will help retailers understand which clustering approaches are most appropriate to employ.

Introduction

Assortment Planning is a hot topic, especially amongst retailers, wholesalers and the software developers that offer solutions to these industries. Yet despite lots of conversation, we hear very little discussion about the various clustering methodologies that lie at the heart of most assortment planning approaches. Parker Avery would like to help remedy that situation by examining the various clustering methodologies that we've encountered through working with a variety of retailers, with the aim of providing some insight into which technique or combination of techniques makes the most sense for your business model. Subsequent Parker Avery publications will address other aspects of clustering and assortment planning.

The Role of Clustering In Assortment Planning

Before we dig into clustering, we need to briefly discuss assortment planning. Assortment planning is a term that has been in widespread use throughout the industry, yet does not have a clear, consistent definition. The meaning can vary depending on the perspective of the user and the situation. The term has been used variously to mean quantifying SKU-level sales and purchases, developing targeted assortments, assortment / space optimization and more. In looking at industry and academic literature dealing with assortment planning ranging over 40 years, nearly every aspect of merchandise planning and space planning has been included. One early attempt at definitive work on the subject¹ included the design of the product hierarchy and layout of the display space as part of the scope of assortment planning. Clearly, that definition is too broad for this current exercise.

For purposes of this conversation, we will define **assortment planning** as “*the practice of developing different assortments for targeted groups of customers.*” There still may be other functions of the assortment plan. It may, for instance, be used to quantify purchases for each item or help determine the amounts of inventory to be distributed to each store and held back for direct sales. Yet, for this discussion the primary purpose of assortment planning is the development of tailored assortments.

Following this definition, clustering is the mechanism that is used to develop those targeted groups of customers. The ideal state of assortment planning would allow the targeting of a collection of products to each individual customer, based on his or her particular preferences. We may eventually be able to deliver on this ideal state through

¹Merchandise Assortment Planning by Charles G. Taylor, 1970, National Retail Merchants Association

digital channels, but for the foreseeable future it will not be attainable in the bricks-and-mortar or catalog channels. This is because a multitude of customers and customer types patronize any individual store location, making individual targeting impossible. Clustering seeks to overcome this challenge by grouping together sales outlets (stores, website, catalogs recipients, etc.) that demonstrate similarities in customer shopping behavior.

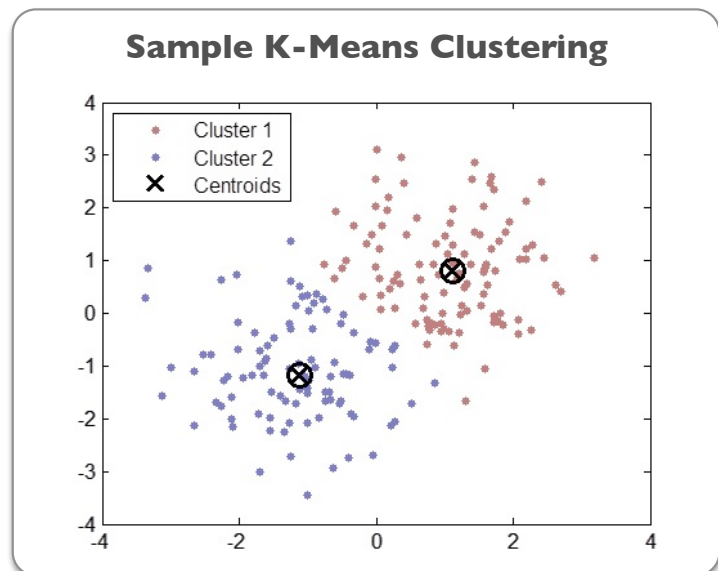
Clustering Defined

We can now turn our attention more fully to the topic of clustering. The term **clustering** refers to *“the process of grouping sales outlets together based on similarities or patterns in their underlying customers’ behavior.”* These similarities are most often gleaned from data related to historic or forecasted sales, or information that is descriptive of the customers or the store. Examples of the latter include demographic or climatic information. Clustering is frequently accomplished using a set of statistical algorithms that assemble a set of objects in such a way that objects in the same group (called a cluster) are more similar to each other than to those in other groups.

The most frequently used statistical method for developing clusters is **K-Means clustering**, which requires the user to specify

a target number of clusters. The algorithm then creates the specified number of groupings, such that the statistical distance between the clusters is maximized. This procedure can be done multiple times and the results compared to determine the optimal number of clusters for the user’s purpose. For retail applications, clusters are typically formed by grouping stores and other sales outlets (such as a website or catalog).

Clustering methodologies have a number of applications in retail merchandising. Statistical grouping of sales outlets can be very useful for allocation, macro-space planning / space brokering, size optimization, determining pricing zones, etc. For the rest of this conversation, though, we will concentrate on the application of clustering to assortment planning activities.



Assortment Clustering vs. Customer Segmentation

In our discussions with clients, we sometimes encounter confusion between assortment clusters and customer segments. Customer segmentation involves the division of a customer base into groups that are similar in ways that are more applicable to product development and marketing. Segmentation uses factors such as age, gender, interests, attitudes and spending habits to classify customers into behavioral or psychographic groups. Examples of these groups might be “Tech-Savvy Millennials” or “Golden Age RV Enthusiasts.” While there are many similarities between assortment clustering and segmentation approaches, most attempts to use customer segments for assortment planning don’t succeed. This is due to the fact that customer segments do not map cleanly onto sales outlets. Any given store will have some representation of most or all customer segments within its customer base. While it is possible to construct assortments based on customer segments, it is very difficult to determine how to assign those assortments to sales outlets. As a rule of thumb, sales outlets should be clustered for assortment planning, while customers should be segmented for product development and marketing purposes.

Clustering Tools

Retailers use a variety of tools to create clusters for assortment planning and other uses. These tools include reports and spreadsheets, specialty statistical analysis software packages (such as SAS and Minitab), clustering solutions tailored specifically for use by retailers, and clustering capabilities that are integrated into a broader assortment planning solution. Depending on the organization’s clustering and assortment planning philosophy, any of these approaches can work well. When considering the best toolset to deploy at your organization, it’s important to consider integration and availability of clustering metadata. Once clusters are created, the cluster assignments typically need to be made available to an assortment planning tool. Your clustering methodology should allow for easy integration of cluster data between your clustering and assortment planning tools. Also, the characteristics of sales outlets that have been placed in the same cluster provide important clues about the product preferences of the underlying customer base to the merchant or assortment planner as they undertake the development of targeted assortments. This cluster

“When considering the best toolset to deploy at your organization, it’s important to consider integration and availability of clustering metadata.”

metadata may communicate basic information such as the number of stores or geographic location of sales outlets. It may also convey more complex insights, such as demographic information or product preferences of the cluster's customer base. Your toolset should be capable of presenting this type of characteristic cluster information to end users as they make product decisions.

Assortment Clustering Approaches

There are many approaches to assortment clustering in use today; some approaches are quite basic, while others require advanced statistical analysis capability. The approaches also may be mixed and matched to meet the particular assortment targeting needs of your organization. We will describe the major ones in this section. One major consideration in determining the ideal clustering method for your company is the complexity that is added to the merchandising process. Some of these approaches require very little ongoing maintenance, while others demand that new clusters be created for each collection or floorset. Some basic methods use the same cluster structure for all categories of product. More complex approaches necessitate the development of different clusters for each category or class of products.



Single Assortment

Each sales outlet receives the exact same selection of items in the assortment

This approach may work well for companies that have a focused, concise product offering that clearly represents the brand image. It also may be applicable to retailers with few outlets or outlets that are situated in very similar markets. Certain premium brands, such as Prada or even Apple, may thrive with this strategy.

On the other hand, retailers with broader product offerings and more diverse store bases may have great difficulty in maximizing sales and margin with this approach. We have frequently heard the lament, "How can we manage multiple assortments when we can't get one right?" The reason is that a single assortment is ill suited to fulfill the needs of a diverse customer and store base.



Channel-Based Clusters

Each channel (bricks and mortar, on-line, catalog, etc.) is a distinct cluster and has its own variant on the assortment

This represents a good preliminary approach to differentiating assortments. It allows the retailer to take advantage of the unique display characteristics of each sales channel, particularly the “endless aisle” offered by websites. It also increases the probability that the retailer can meet a customer’s needs by allowing fulfillment from multiple assortments across multiple channels. Typically, the on-line channel has the broadest offering, with stores and catalogs being culled down from there. Sometimes, retailers will have “retail only” items as well, usually in cases where products are impacted by state regulations (e.g. liquor or firearms) or have physical characteristics that make them impractical to sell on-line.

The downside of this approach is that it can sub-optimize the bricks-and-mortar channel. It is too simplistic to reflect regional and local differences in customer preference and demand. Since it doesn’t allow for the tailoring of assortments within a channel (only across channels), it is likely that retailers following this approach are suffering from slow moving choices in some locations and excess demand in others.



Sales Volume-Based Clusters

Sales outlets are classified based on their historic or forecasted sales volume, expressed in either dollars or units

This is a very common clustering approach, whose main benefit is that it is relatively easy to understand and implement. Frequently, some type of store volume-based attribute is already available in the location data, having been created for use by allocation or replenishment tools. Supporters will use this approach to expand the breadth of the product offering in high volume stores and edit it in low volume stores.

Unfortunately, sales volume-based clustering isn’t very useful for developing differential assortments. Store sales volume typically is driven by population density, traffic patterns, co-tenancy, local competition and other factors not related to the product offering. Stores located in Miami and in Minneapolis coincidentally may be in the same sales volume cluster, but it would be a mistake to assume that they would require the same items. And what to do with high sales volume stores with small selling floors? This approach does not help much in tailoring assortments to those needs.



Store Capacity-Based Clusters

Sales outlets are grouped together based on some measure of their available display space, usually expressed in selling square footage, SKU counts, or on detailed space planning data

This clustering approach is helpful in determining the number of choices to house in each cluster, as it is based on the display capacity of the outlet. It also has the benefit of being relatively easy to understand and execute.

However, it does little to aid in determining how to make up the actual content of those choices – i.e., the assortment of products. As with sales volume-based clusters, stores in Miami and Minneapolis (as an example) may have the same selling square footage, but might require dramatically different assortments. Also, following the pure capacity-based approach does not take into account the sales velocity generated by the outlet. This could result in sending an excessively broad assortment to a large square footage store with poor sales potential.



Sales Volume & Store Capacity-Based Clusters

Sales outlets are clustered together based on a combination of historic or forecasted sales and a measure of capacity

While we mentioned earlier that all of these approaches could be mixed and matched, this particular combination is common enough to merit being discussed separately. It has the benefit of taking into account both capacity and sales volume, so it provides a decent chance of getting the *size* of the assortment correct.

Unfortunately, once again this approach doesn't help much with determining the *content* of the assortment to be assigned to each cluster. Also, this combination of factors has the effect of multiplying the resulting number of clusters, which dramatically increases assortment planning complexity. Most retailers that follow this method end up with a cluster of high capacity / low volume stores, presenting a major challenge:

- Does the retailer provide these stores with an extended assortment to help fill the display space? In so doing, they will be sending many below average performing items to a low volume store, creating markdown jeopardy.
- Or do they send an abbreviated assortment, commensurate with the sales volume, but leave a significant portion of display space empty?

Neither option seems to fit the bill.

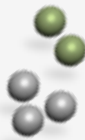


Climate-Based Clusters

Sales outlets are segmented based on seasonal weather patterns

This approach is frequently used by retailers who carry items with pronounced seasonality, such as swimwear, winter coats or patio furniture. While we wouldn't propose that winter boots appropriate for Alaska should be carried in Los Angeles, this method can be less straightforward than it seems. Parker Avery has performed multiple studies on seasonal selling patterns that have shown counter-intuitive results. In one example, a national apparel chain discovered that in January, its bestselling swimsuit stores were located in frigid Minnesota. In another study, no evidence could be found that sales of winter jackets spiked earlier in the North than in the South.

Before embarking on a climate-based clustering effort, we would advise performing in-depth analyses of the regional sales performance of seasonal merchandise to validate the approach. Once the underlying data confirm the validity of a climate-based scheme, these clusters can be used to tailor assortments or adjust the timing of item introductions to closely match local demand patterns.



Store Type-Based Clusters

Sales outlets are grouped based on a salient characteristic of their local market

Store type-based clusters frequently arise from local store managers' or district managers' requests for specific types of merchandise, based on direct customer feedback or perceived market needs. Examples of store type clusters include "campus stores" (that require more back-to-school items and appropriate team merchandise) or "resort stores" (that require beach towels, sunscreen and flip flops throughout the year).

While store types are often identified using input from the store operations organization, this information is sometimes combined with data analysis. Store type clusters tend to be created and maintained manually as a location attribute, but still may be interfaced into an assortment planning solution to allow visibility and planning by end users. This approach can be very effective at capturing some limited localized demand. On the other hand, the manual nature of this approach usually precludes deploying it on a broad scale. Also, assortment requests from store operations can be based on a few anecdotal customer interactions, which may not be representative of true underlying demand. These cases can result in a lot of effort, but ultimately drive few incremental sales.



Competition-Based Clusters

Sales outlets are grouped based on the presence of specific competitors in their market area or based on a measure of competitive intensity

This method is not broadly used, but may have some application for retailers that face strong, differentiated, regional competitors. As an example, a broad-line mass merchandiser may choose to beef up their assortment of hunting, fishing and camping gear if they compete in a market against an outdoor specialty superstore. Many retailers face a distinct set of competitors for their ecommerce channel, and may elect to use this approach to offer special or extended assortments. This approach does not provide merchants and assortment planners any information about the types of products they should add to or edit from their assortments. Instead, competitive shopping and other forms of research must be used to help determine the optimal product mix. Competition-based clusters are also quite useful for price management (but that's a topic for another day).



Demographics-Based Clusters

Sales outlets are segmented based on statistical data about the characteristics of the shopping population

This approach has some benefit, particularly if products within an assortment have a clear appeal to a particular demographic group, such as with ethnic foods or specialized products for the aged. Clusters are created based on characteristics that might include average age, ethnicity, income level, population density, educational level and others.

There are several challenges with the demographics-based technique, however. The first is that the demographic data associated with any particular store may not actually represent the actual shoppers. Most demographic data that is available to retailers is based on U.S. Census data. It represents the characteristics of the population of a certain radius around the store, typically 5 miles. Unfortunately, the population that shops at a particular store is not necessarily representative of the population surrounding the physical address of the store.

Let's examine a retail outlet located adjacent to Penn Station in New York City. The demographics-based clustering approach would suggest that the population shopping that store would resemble that of Manhattan. Yet, since Penn Station is the terminus of the Long Island Railroad, which carries millions of commuters to the city each year, the stores' actual shoppers might more closely resemble the more middle and working class folks from Nassau and Suffolk counties.

One way around this problem is to make use of data about the actual customers of each outlet. This is sometimes obtained through credit card companies, but this approach only captures credit card customers, who may not be representative of the overall customer base. Increasingly, retailers are relying on Customer Relationship Management (CRM) data from loyalty programs to characterize and cluster sales outlets. This method is promising, but may still exclude a significant portion of shoppers that have not joined the loyalty program.

Most importantly, knowing the underlying demographic characteristics of a group of stores does not mean that a merchant knows how to assort to that customer base. The relationship between product preferences and demographics isn't always obvious. For example, consider markets with a high penetration of Hispanic people. The product preferences of such a market in Miami (with its strong Cuban and South American influence) may be entirely distinct from those of Arizona (much more akin to those of neighboring Mexico). In most cases, this kind of information, while accurate, may be misleading.



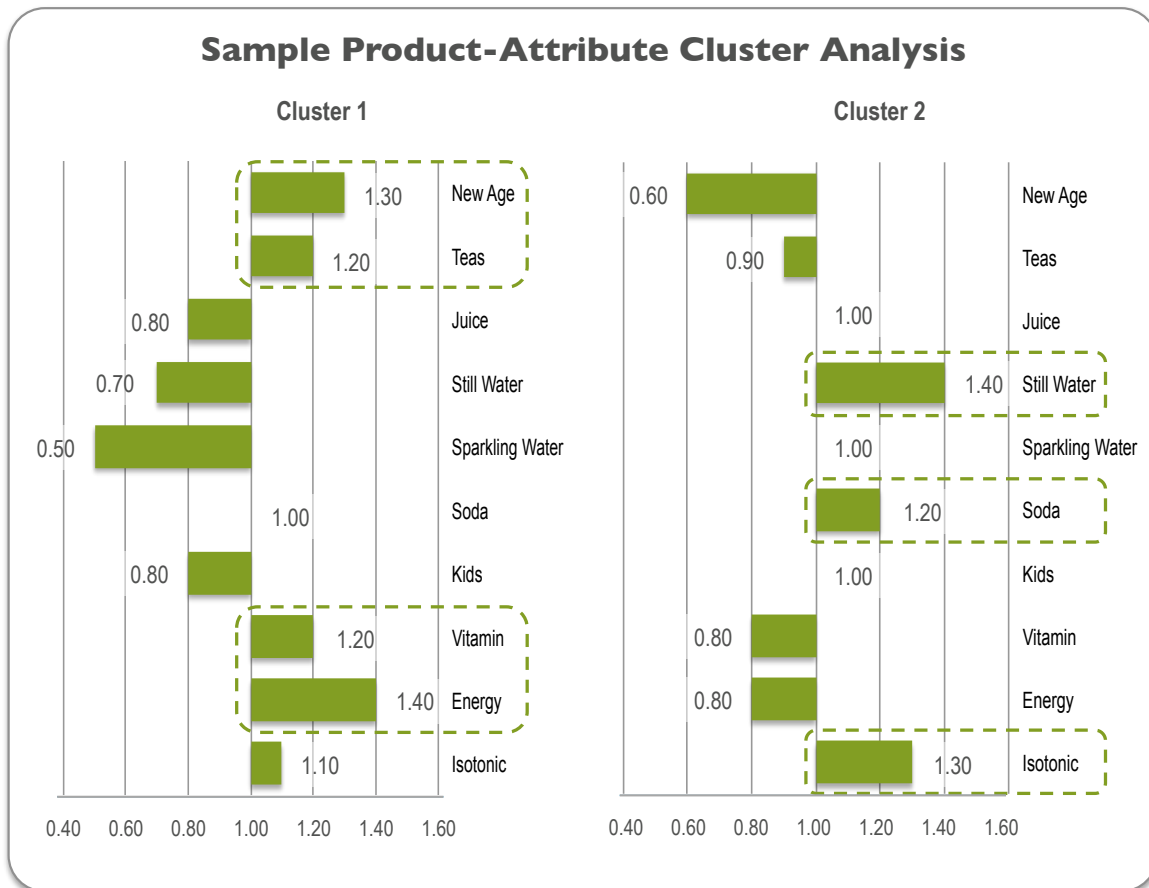
Product Attribute-Based Clusters

Sales outlets are grouped based on sales history of meaningful product attributes of the assortment

In our opinion, this is one of the most valuable clustering approaches and demands a lengthier discussion. This approach has the benefit of the clusters being explicitly tied to the make-up of the assortment. It removes the guesswork on the part of the merchant about which products satisfy the customers in which clusters. To illustrate, an attribute that may be useful for jewelry might be “Material,” with distinct values including Gold, Silver, Stainless Steel, Platinum, Hematite, etc. Outlets are clustered based on their relative sales of products exhibiting these attributes. Should a store exhibit an affinity for silver jewelry (perhaps because it is based in the Southwestern United States), then the merchant can simply assign more silver items to the assortment for that store.

Multiple attributes can be used to describe the same assortment; for example, jewelry could also be described by “Price Point” or “Gem Type.” To identify the best attributes, we recommend undertaking a statistical analysis of the relationship between the available product attributes and sales to determine which attributes drive differential sales performance from outlet to outlet. The attributes with the most impact should be the ones used for clustering.

To further illustrate, let's examine a real world example from the category of "Beverages." In this case the most sales-impactful attribute happened to be "End Use," with attribute values including Isotonic, Energy, Vitamin, Tea, Kid's, Soda, Sparkling Water, etc. Stores were clustered based on the penetration of each of these attributes in their sales history. Below are graphical representations of the penetration of each of those attributes values in two of the resulting clusters.



As you can see, customers in the stores that make up Cluster 1 have a clear preference for New Age, Teas, Vitamin, and Energy drinks. These same customers are not as interested in traditional beverages, such as Still Water, Sparkling Water and Juice. Cluster 2 seems much more oriented toward thirst quenching, over-indexing on Still Water, Soda and Isotonic (such as Gatorade), at the expense of New Age, Vitamin and Energy. The strength of those preferences can even be gauged by the magnitude of the index number. In Cluster 1, customers have bought 1.4X the overall average amount of Energy drinks, while they have purchased half the overall average amount of Sparkling Water. This kind of precise preference data can directly inform the number of choices assigned to each cluster that bear each attribute.

Once attribute clusters are formed, demographic data for each cluster can be analyzed to determine if there are any significant relationships between population characteristics

and cluster membership. If such relationships exist, there is now some compelling insight into the make-up of the customer base of that cluster. If demographics reveal no significant population characteristics for the cluster, all of the necessary information still exists to make intelligent assortment decisions. Demographic cluster characteristics can also be used to create a model to predict which cluster a new store might fit into.

One significant drawback of this approach is that it demands the use of different clusters for each product category, which tends to increase the complexity of creating and maintaining store clusters – especially when faced with a rapidly changing store base. Another potential source of complexity comes with categories that have many different seasonal assortments, for example in apparel. If an apparel retailer has six seasons or collections a year and drops six distinct assortments with different attributes, then the clustering and assortment planning processes may have to be performed six different times.

Another shortcoming of this method is that it does not take into account the display capacity of the stores within each cluster. To overcome this problem, a hybrid method could be employed that includes both the penetration of product attributes and the capacity of the store. Perhaps a preferable method would use attribute-based clustering to determine the content of a “master assortment” for each category. Items within that “master assortment” could then be ranked by sales importance and culled down to fit the available display space in each store.

Summary of Clustering Approaches

Clustering Approach	Single Assortment	Channel-Based	Sales Volume-Based	Store Capacity-Based	Sales Volume and Store Capacity-Based	Climate-Based	Store Type-Based	Competition-Based	Demographics-Based	Product Attribute-Based
Description	Each sales outlet receives the exact same selection of items	Each channel (bricks and mortar, on-line, catalog, etc.) is a distinct cluster	Sales outlets are classified based on their historic or forecasted sales volume	Sales outlets are grouped together based on some measure of their available display space	Sales outlets are clustered together based on a combination of historic or forecasted sales and a measure of capacity	Sales outlets are segmented based on seasonal weather patterns	Sales outlets are grouped based on a salient characteristic of their local market	Sales outlets are grouped based on the presence of specific competitors in their market	Sales outlets are segmented based on statistical data about the characteristics of the shopping population	Sales outlets are grouped based on sales history of meaningful product attributes of the assortment
Advantages	Easy to understand and execute	Can take advantage of “endless aisle” for direct channel and can respond to different competitors by channel	Easy to understand and execute	Explicitly accounts for the capacity of stores in each cluster so the approach helps get the size of the assortment right	Accounts for both sales volume and store capacity, so the approach helps get the size of the assortment right	Supports targeting of items with seasonal or climate-based selling characteristics	Supports targeting of items with local market- based demand characteristics	Supports targeting of assortments to better compete with local competitors	Can target product assortments to different customer bases, based on demographic data	Best approach for targeting assortments to localized product needs
Disadvantages	Does not account for local differences in customer product preferences	Does not allow for tailoring of assortments within the “bricks and mortar” channel	Does not support the determination of the size or content of assortments	Does not support the determination of the content of assortments	Does not support the determination of the content of assortments	Often deployed without sufficient analysis to confirm validity of approach	Typically a manual process, which doesn’t scale well and is subject to anecdotal evidence	Does not help determine the content of assortments targeted against competitors	Links between product preferences and demographic information may not be clear	Adds complexity by requiring multiple clusters that change periodically, does not account for display capacity
When to Use	If you have a focused, concise product offering that clearly represents the brand image or have few outlets or outlets that are situated in very similar markets	If you have distinct on-line competitors or want to take advantage of “endless aisle” for eCommerce, but don’t wish to manage multiple assortments within a channel	Not recommended for assortment planning use	If you have a homogenous customer base with few differences in product preference, but distinct differences in store size and configuration	If you have a homogenous customer base with few differences in product preference, but distinct differences in store size and configuration	If data analysis supports regional or timing differences in sales of seasonal products	If local product needs are clear and confined to a small group of products and / or a small group of sales outlets	If you face strong, differentiated, regional competitors	If products within an assortment have a clear appeal to a particular demographic group	If data analysis demonstrates strong, differential customer preferences for one set of product attributes over another

Final Word

As we have illustrated, clustering for assortment planning is a complex undertaking with many different factors to take into account. Simpler, more straightforward methods tend to have significant shortcomings for creating meaningful differentiated assortments. More sophisticated approaches bring with them increased complexity, which may require more manpower or systems resources to successfully employ. Yet, the financial rewards for getting targeted assortments right are significant. After all, every markdown dollar saved goes directly to the bottom line. It is seductive to gloss over clustering capability when creating an assortment planning strategy or selecting and implementing an assortment planning solution, as clustering is commonplace, highly analytical and too often taken for granted. Don't fall into this trap. Start by clearly defining the purpose and objectives of your assortment planning efforts. Once these are established, the best clustering approach can be identified and operationalized.

The Parker Avery Group

The Parker Avery Group is a boutique strategy and management consulting firm that is a trusted advisor to leading retail brands. We combine practical industry experience with proven consulting methodology to deliver measurable results. We specialize in merchandising, supply chain and the omnichannel business model, integrating customer insights and the digital retail experience with strategy and operational improvements. Parker Avery helps clients develop enhanced business strategies, design improved processes and execute global business models.

Learn more about us at:

www.ParkerAvery.com

for more details
contact:

Clay Parnell

President & Managing Partner | clay.parnell@parkeravery.com

Joshua Pollack

Associate Partner | josh.pollack@parkeravery.com

770.882.2205