Market Segmentation Analysis

**Contributor:**

Karakavalasa venkata pranay

**Date 20/08/2023**

# Abstract:

Market segmentation analysis is the process of dividing a larger market into smaller groups or segments of consumers with similar needs, preferences, behaviors, or characteristics. The purpose of this analysis is to better understand and target specific groups of customers with tailored marketing strategies that meet their specific needs and preferences.

**Step1: Deciding (not) to Segment (Identify the Market)**

**Section 1**: Implications of Committing to Market Segmentation:

* Market segmentation requires a long-term organizational commitment and willingness to make substantial changes and investments.
* There are costs associated with market segmentation such as research, fielding surveys and focus groups, designing multiple packages and advertisements, and communication messages.
* Changes may include developing new products, modifying existing products, changing pricing and distribution channels, and adjusting the internal structure of the organization to target different market segments.

**Section 2**: Implementation Barriers

2.1: The first group of barriers is related to senior management:

* Lack of leadership
* Pro-active championing
* Commitment and involvement in market segmentation process by senior leadership
* Not making enough resources available for market segmentation analysis and implementation

2.2: The second group of barriers is related to organizational culture:

* Lack of market or consumer orientation
* Resistance to change and new ideas
* Lack of creative thinking
* Bad communication and lack of sharing of information and insights across organizational units
* Short-term thinking and unwillingness to make changes
* Lack of training

The lack of a formal marketing function or at least a qualified marketing expert in the organization can also represent major stumbling blocks.

The lack of a qualified data manager and analyst in the organization can be a major obstacle.

Most of these barriers can be identified from the outset of a market segmentation study and proactively removed.

If barriers cannot be removed, seriously consider abandoning the attempt of exploring market segmentation as a potential future strategy.

**Step 2: Specifying the Ideal Target Segment**

**Segment Evaluation Criteria:**

In the process of market segmentation analysis, it is important to consider user input beyond just a briefing at the start or the development of a marketing mix at the end. The segmentation team should select the criteria they want to use to determine how attractive potential target segments are, and assess the relative importance of each criterion to the organization. This results in a diverse set of attractiveness criteria, from which the team can select approximately six segment evaluation criteria with a weight attached to each to indicate their importance to the organization.

# Knock-Out Criteria

Knock-out criteria are used to eliminate market segments that do not qualify for further assessment using segment attractiveness criteria.

The six knock-out criteria are:

1. Homogeneity refers to the similarity among members of a segment.
2. Distinctiveness means that members of a segment should be distinctly different from members of other segments.
3. The segment must be large enough to make it worthwhile to customize the marketing mix.
4. The organization must have the capability to satisfy segment members' needs, which refers to the match between segment needs and the organization's strengths.
5. Identifiability means that segment members can be identified in the marketplace.
6. Reachability means that there must be a way to get in touch with segment members to make the customized marketing mix accessible to them.

Knock-out criteria must be understood by senior management, the segmentation team, and the advisory committee.

**Attractiveness criteria**

* Attractiveness criteria are used to evaluate how attractive potential target segments are.
* Each market segment is rated based on how well it meets each attractiveness criterion.
* Attractiveness criteria are not binary and segments are not assessed as complying or not complying.
* The market segmentation team should have a list of approximately six attractiveness criteria.
* Each attractiveness criterion should have a weight attached to indicate its relative importance to the organization.

# Step 3: Collecting Data

Market segmentation requires the use of segmentation variables and criteria to divide a sample population into different segments. Data can be collected from various sources, including surveys, internal sources, and experimental studies. However, biases may affect the quality of the results obtained from survey data.

**Segmentation Variables:**

* Refers to the variable used to divide a sample into market segments
* Typically a single characteristic of the consumers in the sample

**Segmentation Criteria:**

* Broad term that refers to the nature of the information used for market segmentation
* May relate to specific constructs, such as benefits sought

**Data Collection:**

* Requires identifying the necessary information for informed decisions about market segmentation
* Includes demographic, psychographic, geographic, and behavioral data.

**Survey Data:**

* Cheap and easy to collect, making it a common approach for market segmentation analysis
* Prone to biases that may affect the quality of the results obtained
* Key aspects to consider include the choice of variables, response options, response styles, and sample size.

**Internal Data:**

* Increasingly available and useful for market segmentation analysis
* Examples include scanner data for grocery stores, booking data from airline loyalty programs, and online purchased data.

**Experimental Data:**

* Results from field or laboratory experiments
* Aim to test how people respond to specific stimuli, such as advertisements

**Step 5: Extracting Segments**

**5.1 Grouping Consumers**

Market segmentation is a process of dividing a market into smaller groups of consumers who share similar needs or characteristics. Grouping consumers helps companies tailor their marketing strategies and products to specific segments, resulting in more effective marketing efforts and increased profitability.

However, grouping consumers is not a simple task as consumer data sets are typically unstructured and not well separated. Therefore, market segmentation analysis is exploratory by nature and strongly depends on the assumptions made on the structure of the segments implied by the method.

One of the most popular methods used for market segmentation is cluster analysis, where market segments correspond to clusters. Different algorithms for cluster analysis have different tendencies of imposing structure on the extracted segments. For example, k-means clustering aims at finding compact clusters covering a similar range in all dimensions, whereas single linkage hierarchical clustering constructs snake-shaped clusters.

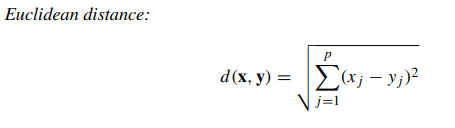
There is no single best algorithm for all data sets, and the choice of algorithm will depend on the data set characteristics and expected or desired segment characteristics. Investigating and comparing alternative segmentation solutions is critical to arriving at a good final solution.

In addition to distance-based and model-based methods for market segmentation, some methods perform variable selection during the extraction of market segments. However, each method has its advantages and disadvantages, and no one method outperforms others in all situations.

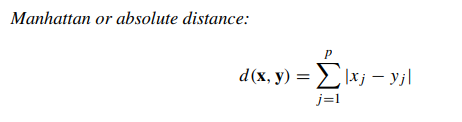
**5.2 Distance-Based Methods**

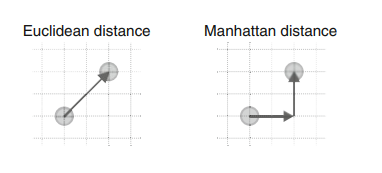
There are several distance measures that can be used to calculate similarity or dissimilaritybetween groups of tourists based on their vacation activity patterns. Here are some commonly used ones:

**Euclidean distance**: This is the most commonly used distance measure, and is defined as the square root of the sum of the squared differences between each pair of activity percentages. It assumes that the variables are continuous and follow a normal distribution.



**Manhattan distance:** This is also known as the L1 distance and is calculated as the sum of the absolute differences between each pair of activity percentages. It is appropriate when the variables are not normally distributed and have outliers.





**Cosine distance:** This is a similarity measure that calculates the cosine of the angle between two vectors of activity percentages. It is useful when the magnitude of the vectors is not important and only their orientation matters.

**Jaccard distance**: This is a distance measure that calculates the dissimilarity between two sets of activity percentages. It is defined as the ratio of the number of elements that are different between the two sets to the total number of elements in the two sets.

Depending on the specific needs of the analysis and the characteristics of the data, one of these distance measures or a combination of them may be used to group tourists into segments based on their vacation activity patterns.

**Hierarchical Methods**

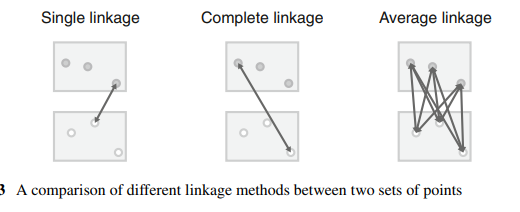
Hierarchical clustering methods group data into segments and are intuitive.

Divisive clustering starts with the complete data set and splits it into two segments, while agglomerative clustering starts with each consumer representing their own segment and merges the closest two segments step-by-step.

Both approaches result in a sequence of nested partitions ranging from partitions containing only one group to n groups.

The linkage method generalizes how distances between groups of observations are obtained.

The standard linkage methods available in the R function hclust() are single linkage, complete linkage, and average linkage.



Different combinations of distance measure and linkage method can reveal different features of the data.

Single linkage is capable of revealing non-convex, non-linear structures, while average and complete linkage extract more compact clusters.

Ward clustering is a popular alternative method based on squared Euclidean distances.

The result of hierarchical clustering is typically presented as a dendrogram, which is a tree diagram showing the sequence of nested partitions.

**Partitioning Methods**

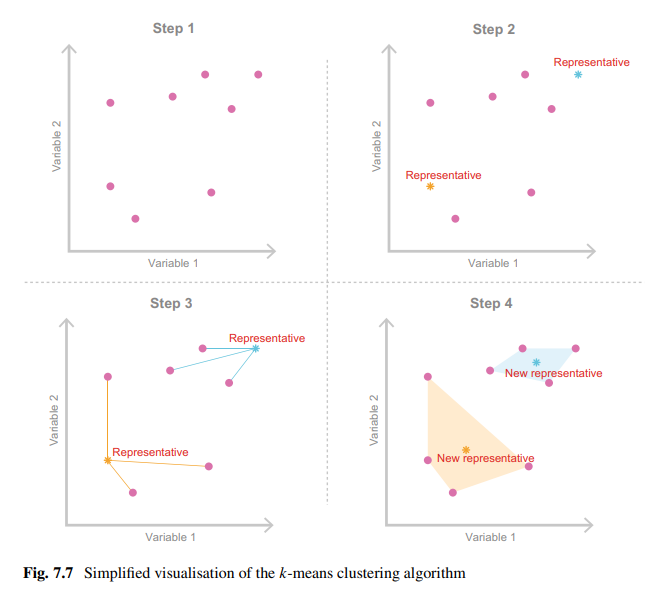
* Hierarchical clustering is best for small datasets with up to a few hundred observations.
* For larger datasets, clustering methods that create a single partition are more suitable.
* Instead of computing all pairwise distances between observations, distances between each observation and the center of segments can be computed.
* For a dataset with 1000 consumers, agglomerative hierarchical clustering would have to calculate 499,500 distances for the pairwise distance matrix between all consumers.
* Partitioning clustering algorithms that aim to extract a specific number of segments only have to calculate between 5 and 5000 distances at each step.
* It's better to optimize specifically for extracting a few segments rather than building the complete dendrogram and then heuristically cutting it into segments.

1. **k-Means and k-Centroid Clustering**

K-means clustering is a partitioning method in unsupervised machine learning used for dividing a dataset into groups (clusters) based on similarity/dissimilarity. The objective is to create clusters of data points in a way that points within the same cluster are as similar as possible to each other, while points in different clusters are as dissimilar as possible.

The method involves the following steps:

1. Specify the desired number of clusters (k).
2. Randomly select k observations (data points) from the dataset and use them as the initial set of cluster centroids.
3. Assign each observation to the nearest cluster centroid based on the chosen distance measure (usually squared Euclidean distance).
4. Recalculate the cluster centroids as the mean (in the case of squared Euclidean distance) of all observations assigned to that cluster.
5. Repeat steps 3 and 4 until convergence is achieved (i.e., no more updates are needed or a maximum number of iterations has been reached).



K-means clustering is an iterative algorithm that aims to minimize the sum of squared distances between each data point and its assigned cluster centroid. The algorithm is not guaranteed to find the global optimum, but it usually converges to a local minimum, which is a suboptimal solution. Additionally, there are variations of the k-means algorithm, such as k-medians and k-modes, that use different distance measures and methods for calculating the cluster centroids.

1. **“Improved” k-Means**

On improving the k-means clustering algorithm. Using "smart" starting values rather than randomly drawing k consumers from the data set can certainly help avoid the problem of the algorithm getting stuck in a local optimum. It's interesting to note that using starting points that are evenly spread across the entire data space can better represent the entire data set and potentially lead to better solutions.

1. **Hard Competitive Learning**

Hard competitive learning is also known as learning vector quantisation and differs from k-means in how segments are extracted.

Both methods minimize the sum of distances from each consumer to their closest representative (centroid), but the process is different.

K-means uses all consumers in the data set at each iteration to determine new centroids, while hard competitive learning randomly picks one consumer and moves its closest centroid a small step towards the randomly chosen consumer.

Different segmentation solutions can emerge from the two methods even if the same starting points are used.

Hard competitive learning may find the globally optimal solution while k-means gets stuck in a local optimum, or vice versa.

Neither method is superior to the other; they are just different.

Hard competitive learning has been used in market segmentation analysis for segment-specific market basket analysis.

Hard competitive learning can be computed in R using the cclust function from the flexclust package, with the method parameter set to "hardcl".

1. **Neural Gas and Topology Representing Networks**

Neural gas algorithm is a variation of hard competitive learning that adjusts the location of the second closest segment representative towards the randomly selected consumer.

Topology representing networks (TRN) extends the neural gas algorithm by building a virtual map where similar segment representatives are placed next to each other.

The segment neighborhood graph can be generated from the final segmentation solution of any clustering algorithm by counting how many consumers have certain representatives as closest and second closest.

There is currently no implementation of the original TRN algorithm in R, but using neural gas in combination with neighborhood graphs achieves similar results.

Different segmentation solutions can emerge from different clustering algorithms, including k-means, hard competitive learning, neural gas, and TRN.

Having a larger toolbox of algorithms available for exploration is of great value in data-driven market segmentation analysis.

1. **Self-Organising Maps**

Self-organizing maps (SOMs) are a variation of hard competitive learning used for market segmentation.

SOMs position segment representatives (centroids) on a regular grid.

The algorithm is similar to hard competitive learning, where a single random consumer is selected and the closest representative moves towards it.

Representatives that are direct grid neighbors of the closest representative also move towards the random consumer.

The adjustments to the locations of the centroids get smaller and smaller until a final solution is reached.

The advantage of SOMs is that the numbering of market segments aligns with the grid along which all segment representatives are positioned.

However, the sum of distances between segment members and segment representatives can be larger than for other clustering algorithms due to the restrictions imposed by the grid.

Comparisons of SOMs and topology representing networks with other clustering algorithms are provided in literature.

1. **Neural Networks**

Auto-encoding neural networks provide a different approach to cluster analysis compared to traditional clustering methods. The key idea behind auto-encoders is to use a neural network to learn a compressed representation of the input data, such that the compressed representation can be used as a representation of clusters or segments. The process of learning this compressed representation is often referred to as training the network.

The architecture of an auto-encoder typically consists of an input layer, a hidden layer, and an output layer. The input layer takes the raw data as input, the output layer produces a reconstructed version of the input data, and the hidden layer provides a compressed representation of the input data. During training, the network is optimized to minimize the difference between the input data and the reconstructed output data.

Once the network is trained, the hidden layer can be used as a representation of clusters or segments. Consumers that have similar hidden layer values are considered to be members of the same segment. Auto-encoder clustering typically results in fuzzy segmentations, where consumers may belong to multiple segments with membership values between 0 and 1.

Auto-encoding neural networks have the advantage that they can learn non-linear relationships between input variables, which traditional clustering methods may not be able to capture. Additionally, auto-encoders can learn to represent data in a lower-dimensional space, which can be useful for data visualization.

There are several implementations of auto-encoding neural networks available in popular programming languages such as R and Python. The R package autoencoder provides an implementation of auto-encoding neural networks, while the fclust package provides implementations of other fuzzy clustering algorithms.

**Hybrid Approaches**

1. **Two-Step Clustering**

Two-step clustering is a data clustering technique that involves two steps, as the name suggests. In the first step, a partitioning clustering method, such as k-means, is used to divide the data into a large number of small, homogeneous clusters. The primary objective of this step is to reduce the size of the data set by retaining only one representative member of each cluster. This step is also referred to as vector quantization. In the second step, a hierarchical clustering method is applied to the representative members obtained in the first step, and the original data is linked to the resulting segmentation solution.

The second step uses the cluster centers and segment sizes obtained from the first step as input to the hierarchical clustering method. The resulting dendrogram produced by hierarchical clustering is analyzed to identify the natural segments within the data. However, it cannot be determined which observation belongs to which segment without linking the original data with the hierarchical clustering solution. This is done using the twoStep() function, which takes as arguments the hierarchical clustering solution, the cluster memberships of the original data obtained with the partitioning clustering method, and the number of segments to extract.

Two-step clustering is often used in situations where the number of natural segments in the data is unknown or not well defined. The two-step approach allows for a more robust and accurate segmentation solution by combining the strengths of both partitioning and hierarchical clustering methods. The approach has been applied in various fields, including market research, social science, and healthcare.

1. **Bagged Clustering**

Bagged clustering is a type of clustering algorithm that combines partitioning clustering and hierarchical clustering techniques while also using bootstrapping. Bootstrapping is a process of randomly drawing samples from the original data set, with replacement. The main advantage of this method is that it makes the segmentation solution less dependent on the exact people contained in consumer data.

In the first step of bagged clustering, the data set is bootstrapped to create many random samples. For each sample, a partitioning algorithm is applied to cluster the data, and the resulting centroids are saved. These centroids are then used as the data set for hierarchical clustering. The dendrogram from hierarchical clustering provides clues about the best number of market segments to extract.

Bagged clustering is suitable for situations where niche markets are suspected, standard algorithms might get stuck in bad local solutions, or hierarchical clustering is preferred, but the data set is too large. It consists of five steps, including creating bootstrap samples, repeating the partitioning method, using cluster centers to create a new data set, calculating hierarchical clustering, and determining the final segmentation solution.

Bagged clustering has been applied to tourism data and has been successful in identifying market segments based on winter vacation activities, as illustrated by the winter vacation activities data from the Austrian National Guest Survey.

**5.3 Model-Based Methods**

Model-based methods are a flexible and powerful alternative to distance-based methods for market segmentation analysis. They rely on assumptions about the underlying structure of the market segments, but they allow for more complex and nuanced relationships between consumer characteristics and segment membership. By using a range of extraction methods, data analysts can better understand the nature of the market and make more informed marketing decisions.

A mixture of several multivariatenormal distributions is a popular choice for finite mixture models when the segmentation variables are metric. This is because the multivariate normal distribution can easily model covariance between variables, which is often present in real-world data. As you mentioned, this covariance can arise from physical measurements on humans, such as height, arm length, leg length, or foot length, or from business data, such as prices in markets with many players. By using a mixture of normal distributions, we can segment the data based on the different mean and covariance structures of each segment, which can provide insights into the underlying patterns and relationships in the data

Finite mixtures of regression models can be used to identify these different segments and estimate the regression relationship between the number of rides and the entrance fee separately for each segment. The basic idea behind finite mixture models is to assume that the data is generated by a mixture of several subpopulations, each with its own probability distribution.

In the context of finite mixtures of regression models, we assume that the data is generated by a mixture of several subpopulations, each with its own regression function. The probability of belonging to each subpopulation is represented by a weight. The regression coefficients and the error term of each subpopulation are estimated separately.

By estimating these parameters, we can identify the different subpopulations or segments of consumers with different levels of willingness to pay based on the number of rides. We can also estimate the regression relationship between the number of rides and the entrance fee separately for each segment. This allows us to identify the optimal price for each segment, which can improve revenue and profitability.