Ten Steps of Market Segmentation Analysis

Submitted by-

Keya Suvarna

Shambhavi Parashar

Kongari Aarthi

Kumar Kausik Sahu

Ingole Rohit Rameshrao

Step 1: Deciding (not) to Segment

Deciding whether to pursue market segmentation is a crucial step for organizations. Market segmentation involves dividing a market into distinct groups with similar needs or characteristics. While it can be an effective strategy, it requires long-term commitment and substantial investment. Before proceeding, organizations must understand the implications and potential barriers.

Implications of Committing to Market Segmentation:

- Market segmentation requires long-term commitment and substantial changes within the organization.
- It involves costs such as research, designing products and advertisements, and adjusting communication strategies.
- Organizations should only pursue segmentation if the expected increase in sales justifies the investment.

Implementation Barriers:

- Lack of leadership and commitment from senior management can hinder successful implementation.
- Organizational culture, such as resistance to change and lack of market orientation, can also pose barriers.
- Insufficient training and resources, including a lack of qualified personnel, can impede progress.
- Objective restrictions like financial constraints or the inability to make necessary structural changes can be obstacles.

Step 1 Checklist:

- Assess the organization's market orientation, willingness to change, long-term perspective, openness to new ideas, and communication effectiveness.
- Secure commitment and involvement from senior management.
- Ensure understanding of market segmentation concepts and implications.
- Form a dedicated team with necessary expertise.
- Develop clear objectives, structured processes, and allocate responsibilities.
- Ensure adequate time and resources for the segmentation analysis.

In summary, deciding whether to pursue market segmentation requires careful consideration of organizational readiness, commitment, and potential barriers. It involves assessing internal factors, securing management support, and forming a capable team to navigate the segmentation process effectively.

Step 2: Specifying the Ideal Target Segment

When analysing market segments, it's important to involve users throughout the process, not just at the beginning or end. Organizations must contribute significantly to the analysis, particularly in determining segment evaluation criteria. There are two types of criteria: knock-out criteria and attractiveness criteria.

Knock-out criteria are essential features that segments must have to be considered for targeting. These include homogeneity, distinctiveness, size, matching organizational strengths, identifiability, and reachability.

Attractiveness criteria help evaluate the relative attractiveness of segments that meet the knock-out criteria. These criteria are diverse and include factors like market size, growth, profitability, accessibility, and compatibility with the organization's strengths

Implementing a structured process is crucial for evaluating segments effectively. Using tools like segment evaluation plots can help visualize segment attractiveness and organizational competitiveness. It's essential to involve a diverse team representing various organizational units to ensure comprehensive analysis and stakeholder buy-in.

At the end of this step, the segmentation team should have a list of approximately six attractiveness criteria, each weighted according to its importance to the organization. This list should be discussed with the advisory committee for further refinement and approval.

To sum it up, Step 2 involves defining knock-out and attractiveness criteria, involving stakeholders, and laying the groundwork for data collection and target segment selection.

Step 3: Collecting Data

Market segmentation is about dividing customers into groups based on similarities. There are two main approaches: common sense and data-driven segmentation. In common sense segmentation, we use one simple characteristic, like gender, to split customers into groups. These groups are then described using other characteristics, like age or vacation preferences.

Data-driven segmentation is more complex. It uses multiple characteristics, called segmentation variables, to create segments based on shared traits or behaviours. For example, instead of just gender, we might use vacation preferences to create segments of tourists. Choosing the right segmentation criteria is crucial. It can be based on geography (where people live), socio-demographics (like age or income), psychographics (beliefs and interests), or behaviour (like past purchases).

Each criterion has pros and cons. Geographic segmentation is simple but may not reveal much about consumer preferences. Socio-demographic segmentation is easy to use but may not explain why people buy certain products. Psychographic segmentation looks at beliefs and interests, providing insight into consumer behaviour, but it's complex to implement. Behavioural segmentation is based on actual behaviour, which is very relevant but may not always be easy to collect.

Which tells us that, good data quality is essential for effective segmentation. It helps identify the right criteria and accurately describe customer segments. And when it comes to choosing segmentation criteria, simpler is often better, as long as it works for your product or service.

Choosing the right variables for market segmentation is really important. In simple terms, it means picking the right things to look at when dividing customers into groups. If we include too many unnecessary things, it can make surveys long and tiring for people to fill out. This can make them give lower quality answers.

When collecting data, it's crucial to avoid "noisy" variables, which are things that don't really help us understand customer groups. These can mess up the results and make it hard for algorithms to find the right groups. The way we ask questions in surveys also matters. We should try to keep questions clear and not repeat things unnecessarily. This helps to avoid confusing people and getting wrong answers. The options we give people to answer questions are also important.

Some ways of answering are better than others for analysing the data later. For example, it's easier to use data where people can pick from a range of options or give a number. Sometimes, people answer surveys in a certain way because of habits, not because of what they really think. This can affect the results, so it's important to watch out for these "response styles." Having enough people in our survey is really important too. If we don't have enough, it can be hard to find the right customer groups. We need a good number of people for each thing we're looking at in the survey. Internal data from companies, like what people buy or book online, can be really useful for segmentation. But we have to be careful because it might only show what current customers do, not what other potential customers might do.

Lastly, experimental data from tests or studies can also help with segmentation. These can give us insight into how people respond to different things, like ads or product features, which can be useful for dividing customers into groups based on their preferences.

Step 4: Exploring Data

4.1 A First Glimpse at the Data

- Data exploration helps clean, pre-process data, and choose segmentation methods.
- In that matter in particular we see behavior of data and predict which model is suitable for it.
- It helps identify measurement levels, investigate variable distributions, and assess dependencies between variables.

4.2 Data Cleaning

- Before analysis, it's crucial to ensure all values are recorded correctly and consistently labeled.Metric variables like age are checked for plausible ranges.Categorical variables are checked for consistent levels.
- To find any irregularities or mistakes in the data entry or collection process, summary statistics are analyzed
- Although time-consuming, cleaning data using code ensures full documentation and reproducibility.
- Clicking in a spreadsheet may be error-prone and less reproducible compared to coding the cleaning steps.
- Reproducibility allows for the exact same procedure to be applied when new data is added or when monitoring segmentation solutions over time.

4.3 Descriptive Analysis

- Descriptive numeric and graphic representations provide insights into the data.
- In that graphical representation we use tools like Excel,power book.
- Histograms: Visualize the distribution of numeric variables.
- Box Plots: Illustrate the five-number summary (min, 1st quartile, median, 3rd quartile, max) and identify outliers.Represents the five-number summary of numeric variables.Provides insight into distributional properties assuming unimodality.elps identify outliers beyond 1.5 times the interquartile range. Outliers are depicted as circles outside the whiskers.
- Scatter plots: Display relationships between two numeric variables.
- Bar plots: Present frequency counts for categorical variables.
- Mosaic plots: Show associations among multiple categorical variables.

4.4 Pre-Processing

- Pre-processing alters the data and should be done with careful consideration of the implications.
- It's important to evaluate whether the transformation preserves the original meaning of the variables and is suitable for the intended analysis.
- Strong arguments or evidence supporting the conversion should be considered before applying pre-processing steps.

4.4.1 Categorical Variables

- Two pre-processing procedures are often used for categorical variables. One is merging levels of categorical variables.

1. Merging levels
    - Useful when original categories are too numerous or differentiated.
    - Simplifies analysis by reducing the number of categories.
    - Often applied to variables with many distinct levels, such as income brackets or education levels.
    - Helps improve interpretability and model performance by reducing noise from small categories.

2. categorical variables
    - Applied when converting categorical variables to numeric ones makes sense and preserves the meaning of the data.
    - Ordinal data can sometimes be converted to numeric if distances between categories are approximately equal.
    - Popular agreement scales (e.g., Likert scales) may be treated as numeric if distances between response options are assumed to be equal.
    - Binary categorical variables can be directly converted to numeric variables with 0 and 1.

4.4.2 Numeric Variables

- In distance-based segment extraction methods, the relative importance of a segmentation variable varies depending on its range of values.
- For example, disparities in scale can result in uneven variable weighting in a situation involving binary and continuous data.
- By putting variables on a uniform scale, standardization aids in balancing their influence.
- A standardized variable having a mean of 0 and a standard deviation of 1 is produced by removing the mean and dividing by the standard deviation during the standardization process.
- This guarantees equitable comparison and allocation of weights among the variables during the segmentation procedure, regardless of their initial scales.
- In order to obtain precise and significant segmentation findings, standardization is essential.
- In the presence of outliers or skewed distributions, alternative standardization methods may be preferable.
- Robust estimates for location and spread, such as the median and interquartile range, are more resistant to outliers and provide a more accurate representation of central tendency and variability.

4.5 Principal Components Analysis

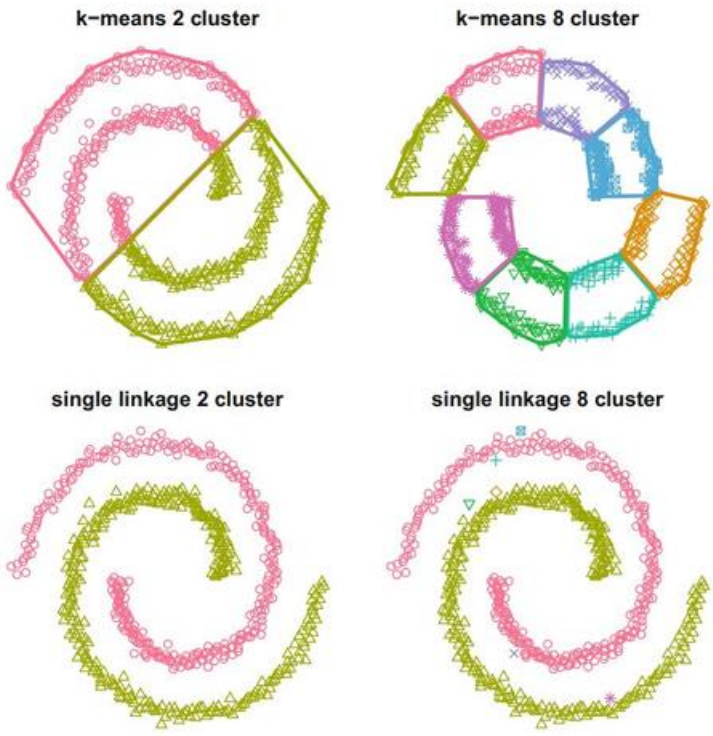
- PCA (Principal Component Analysis) is a mathematical technique that helps in reducing the number of dimensions in a dataset while retaining the maximum amount of information.
- A multivariate dataset with metric variables is transformed into a new dataset with uncorrelated variables called principal components using principal components analysis.

- The significance of these components is arranged as follows: the first component captures the most variability, followed by the second, and so forth.
- PCA adopts a distinct perspective on the data while preserving the relative placements of the observations.
- It operates using the correlation or covariance matrix of numerical variables. It is better to use the correlation matrix if the data ranges are different.
- PCA is frequently used, usually with the first few main components, to project high-dimensional data into lower dimensions for display purposes.
- The principle component standard deviations and a rotation matrix showing the contributions of the original variables to each component are among the outputs of PCA.
- The significance of each primary component is interpreted using summary statistics like cumulative proportion, proportion of explained variance, and standard deviation.
- PCA is helpful for finding strongly correlated variables and for data exploration. By eliminating unnecessary variables from the segmentation base, PCA insights can lower dimensionality without sacrificing the original variables.

Step-5 Extracting Segments

Grouping Consumers: Data-driven market segmentation analysis is exploratory by nature. Consumer data sets are typically not well structured. Consumers come in all shapes and forms. The combination of exploratory methods and unstructured consumer data means that results from any method used to extract market segments from such data will strongly depend on the assumptions made on the structure of the segments implied by the method. Many segmentation methods used to extract market segments are taken from the field of cluster analysis. selecting a suitable clustering method requires matching the data analytic features of the resulting clustering with the context-dependent requirements that are desired by the researcher. It is also important to understand how different algorithms impose structure on the extracted segments.

One of the most illustrative examples of how algorithms impose structure is shown in Fig. 7.1. In this figure, the same data set – containing two spiralling segments – is segmented using two different algorithms, and two different numbers of segments. The top row in Fig. 7.1 shows the market segments obtained when running k-means cluster analysis (for details see Sect. 7.2.3) with 2 (left) and 8 segments (right), respectively



The bottom row in Fig. 7.1 shows the market segments obtained from single linkage hierarchical clustering (for details see Sect. 7.2.2). This algorithm correctly identifies the existing two spiralling segments, even if the incorrect number of segments is specified up front. This is because the single linkage method constructs snake-shaped clusters. When asked to return too many (8) segments, outliers are defined as micro-segments, but the two main spirals are still correctly identified. kmeans cluster analysis fails to identify the spirals because it is designed to construct round, equally sized clusters.

This illustration gives the impression that single linkage clustering is much more powerful, and should be preferred over other approaches of extracting market segments from data. Consider the problem of finding groups of tourists with similar activity patterns when on vacation. A f ictitious data set is shown in Table 7.2. It contains seven people indicating the percentage of time they spend enjoying BEACH, ACTION, and CULTURE when on vacation. Anna and Bill only want to relax on the beach, Frank likes beach and action, Julia and Maria like beach and culture, Michael wants action and a little bit of culture, and Tom does everything. Market segmentation aims at grouping consumers into groups with similar needs or behaviour, in this example: groups of tourists with similar patterns of vacation activities. Anna and Bill have exactly the same profile, and should be in the same segment. Michael is the only one not interested in going to the beach, which differentiates him from the other tourists. In order to find groups of similar tourists one needs a notion of similarity or dissimilarity, mathematically speaking: a distance measure

Data set and segment characteristics informing extraction algorithm selection:

Data set characteristics:

– Size (number of consumers, number of segmentation variables)

– Scale level of segmentation variables (nominal, ordinal, metric, mixed)

– Special structure, additional information.

Segment characteristics:

– Similarities of consumers in the same segment

– Differences between consumers from different segments

– Number and size of segments

The most popular extraction methods used in market segmentation, and point out their specific tendencies of imposing structure on the extracted segments.

Distance-based methods:

Distance-based methods are described first. Distance-based methods use a particular notion of similarity or distance between observations (consumers), and try to find groups of similar observations (market segments). And distance-based methods in machine learning are algorithms that make predictions based on the similarity or dissimilarity between data points. They rely on measuring the distance between data points in a multidimensional space, often using metrics like Euclidean distance or cosine similarity. These methods are commonly used in clustering, classification, and anomaly detection tasks. By identifying patterns or relationships based on proximity, distance-based methods can effectively group similar data points or classify unknown instances. However, they may be sensitive to the choice of distance metric and can suffer from the curse of dimensionality in high dimensional spaces.

Distance measures quantify the similarity between data points, crucial for clustering algorithms like K-means and hierarchical clustering.

Euclidean Distance:

• Euclidean distance is a measure of the straight-line distance between two points in Euclidean space.

• It is calculated as the square root of the sum of squared differences between corresponding coordinates of the points.

• This distance metric is commonly used in machine learning for tasks like clustering, classification, and regression.

• For two points (x1, y1) and (x2, y2) in a 2D space, the Euclidean distance is given by $\sqrt{((x2 - x1)^2 + (y2 - y1)^2)}$. • It represents the shortest path between two points in a geometric sense.

Manhattan Distance:

• Manhattan distance, also known as city block distance or taxicab distance, measures the distance between two points along axes at right angles.

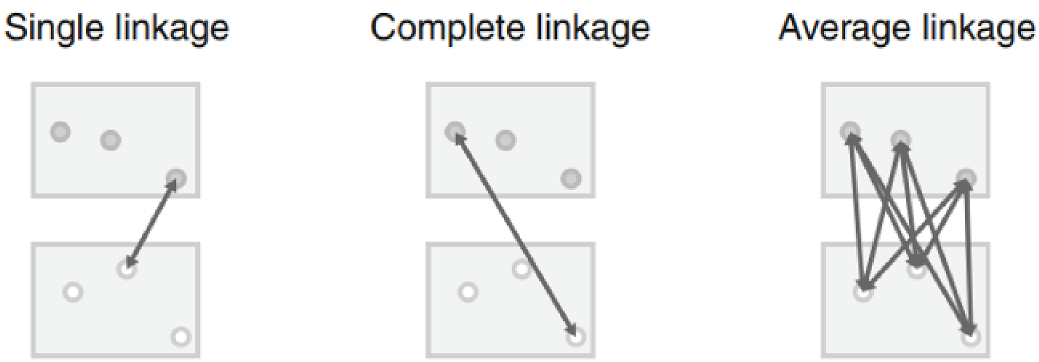
• It is calculated as the sum of the absolute differences between corresponding coordinates of the points.

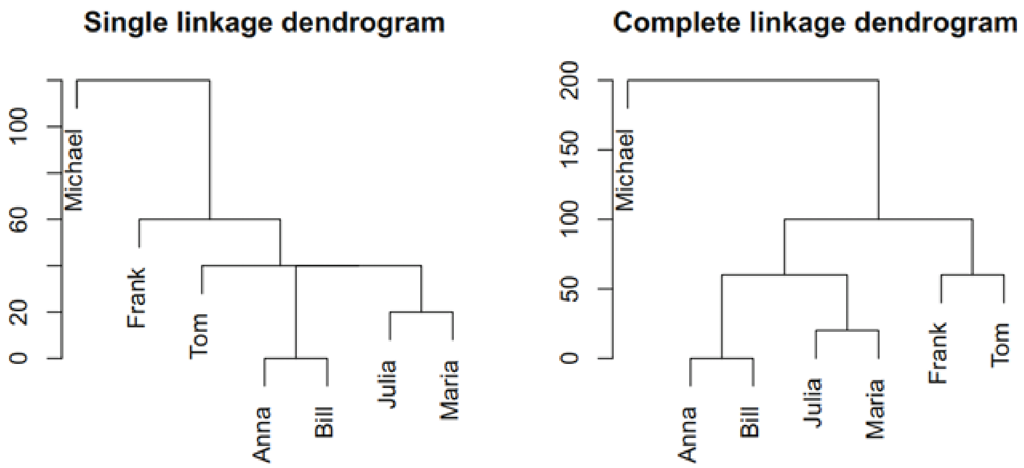
• Manhattan distance is often used in scenarios where movement can only occur along grid-like paths, such as navigation in cities.

• For two points (x1, y1) and (x2, y2) in a 2D space, the Manhattan distance is given by |x2 - x1| + |y2 - y1|.

• Unlike Euclidean distance, Manhattan distance accounts for only horizontal and vertical movements, making it suitable for certain types of data analysis, such as image processing or circuit design.

Hierarchical Methods: Hierarchical methods organize data into a tree-like structure, grouping similar data points at lower levels and forming clusters hierarchically. Hierarchical clustering methods are the most intuitive way of grouping data because they mimic how a human would approach the task of dividing a set of n observations (consumers) into k groups (segments). If the aim is to have one large market segment (k = 1), the only possible solution is one big market segment containing all consumers in data X.





Partitioning Methods

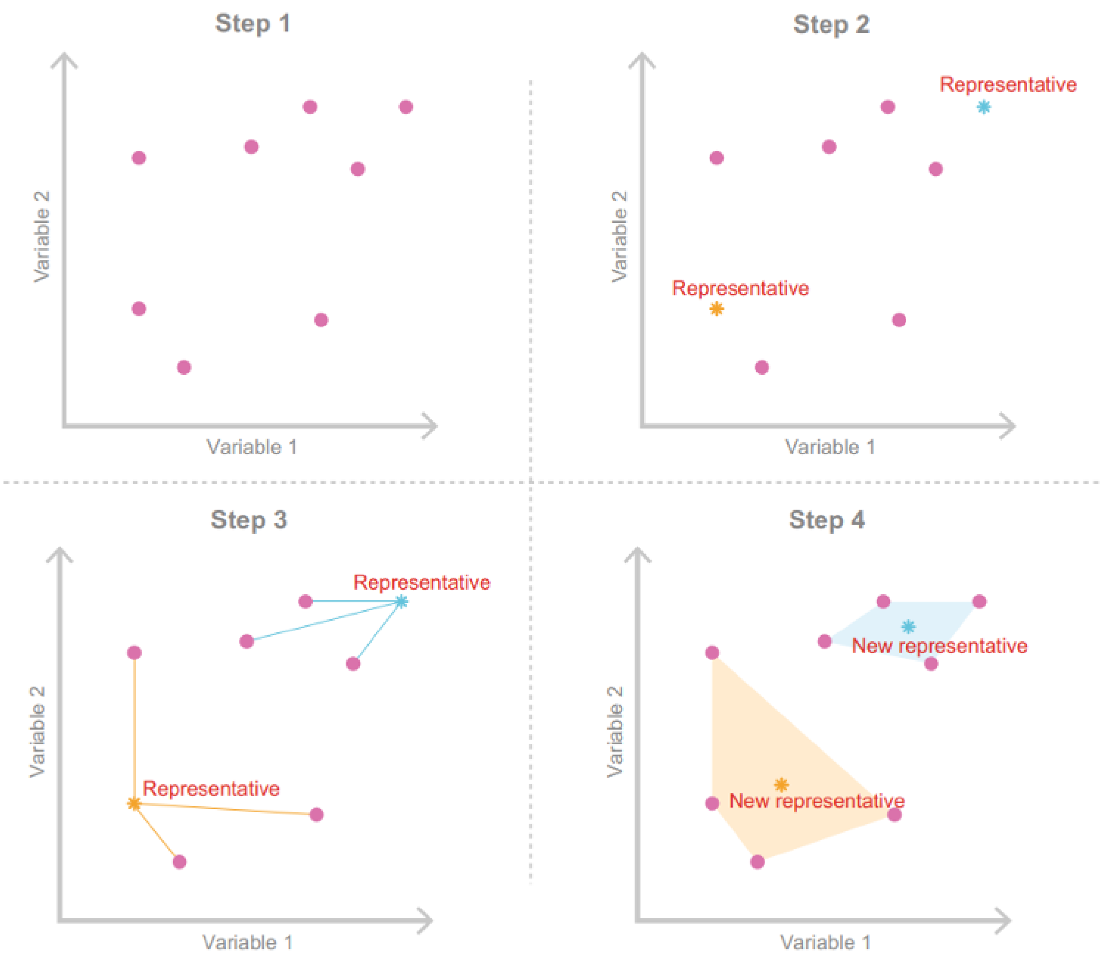
Partitioning methods like K-means partition data into non-overlapping clusters based on distance from centroids. k-Means and k-Centroid Clustering K-means clustering aims to minimize the within-cluster variance by iteratively updating cluster centroids and reassigning data points.

K-centroid clustering extends K-means by allowing each cluster to have multiple centroids, offering more flexibility in representing cluster shapes and sizes.

• Specify the desired number of segments k.

• Randomly select k observations (consumers) from data set X (see Step 2 in Fig. 7.7) and use them as initial set of cluster centroids C = {c1,..., ck}.

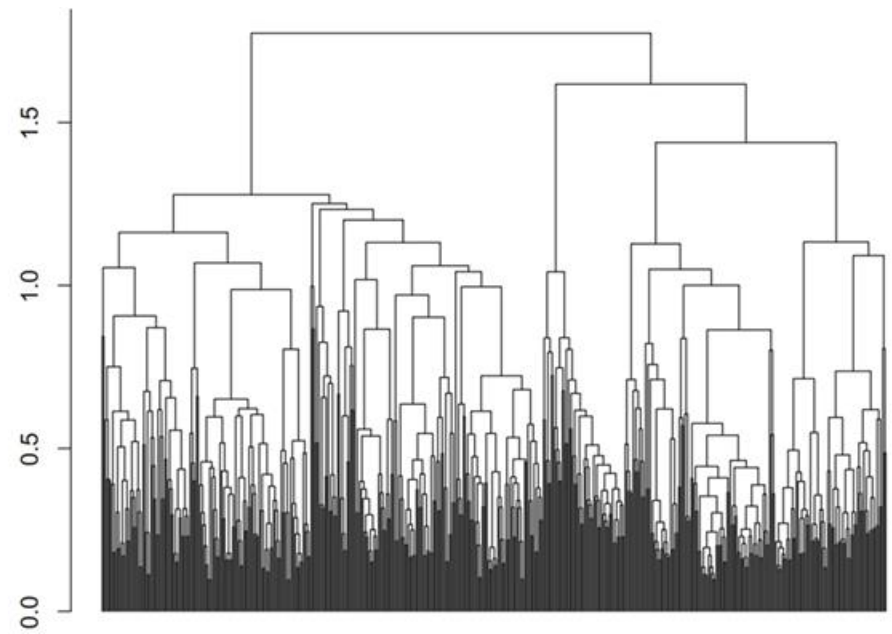
If five market segments are being extracted, then five consumers are randomly drawn from the data set, and declared the representatives of the five market segments.



Hybrid approaches

Hybrid approaches combine multiple machine learning techniques to leverage their respective strengths and improve performance.

Bagged Clustering: Bagging involves training multiple clustering models on different subsets of the data and combining their results to improve accuracy and robustness.



Two-Step Clustering: This method first partitions the data into pre-clusters based on similarity and then refines these clusters using a different algorithm to produce the final clustering solution.

Model-Based Methods:

20 Distance-based methods have a long history of being used in market segmentation analysis. More recently, model-based methods have been proposed as an alternative. Model-based methods in clustering aim to assign data points to clusters based on the assumption that the data are generated from a probabilistic model.

Unlike distance-based methods, which rely on proximity measures, model-based methods explicitly model the probability distribution of the data. One of the most popular model-based clustering algorithms is the Gaussian Mixture Model (GMM). In GMM, each cluster is assumed to follow a Gaussian distribution, characterized by its mean and covariance matrix. The algorithm iteratively estimates the parameters of these Gaussian distributions to maximize the likelihood of the observed data.

Expectation-Maximization (EM) algorithm is commonly used to optimize the parameters. The EM algorithm begins with an initial guess of the parameters and iterates between two steps: the E-step, where it calculates the probabilities of data points belonging to each cluster based on the current parameter estimates, and the M-step, where it updates the parameters based on these probabilities. This process continues until convergence, where the parameter estimates no longer change significantly between iterations. One advantage of model-based methods is their ability to handle clusters of arbitrary shapes and densities.

Unlike distance-based methods, which assume spherical clusters and struggle with non-linear boundaries, model-based methods can capture complex data distributions using flexible probability distributions. Another advantage is the ability to estimate uncertainty in cluster assignments. Since model based methods provide probabilistic assignments, they can quantify the uncertainty associated with each data point's cluster membership. This can be useful in cases where the data points are not clearly separable or when dealing with noisy data.

However, model-based methods also have limitations. They are sensitive to the choice of the underlying probability distribution and the initialization of parameters, which can lead to suboptimal results. Additionally, they are computationally more expensive than distance-based methods, especially when dealing with high-dimensional data or large datasets.

Despite these challenges, model-based methods offer a powerful approach to clustering that can capture complex data structures and provide probabilistic interpretations of cluster assignments. They are particularly useful in scenarios where the data follow distinct statistical distributions and when uncertainty in cluster assignments is of interest.

Finite Mixtures of Distributions: This method involves modeling the data as a combination of multiple probability distributions, allowing for more flexible and complex data representations. It's commonly used in clustering tasks where the data may not fit a single distribution.

Normal Distributions: For metric data, the most popular finite mixture model is a mixture of several multivariate normal distributions. The multivariate normal distribution can easily model covariance between variables; and approximate multivariate normal distributions occur in both biology and business. For example, physical measurements on humans like height, arm length, leg length or foot length are almost perfectly modelled by a multivariate normal distribution. All these variables have an 21 approximate univariate normal distribution individually, but are not independent of each other. Taller people have longer arms, longer legs and bigger feet. All measurements are positively correlated

Binary Distributions: Binary distributions are probability distributions that model binary outcomes, where the outcome can take only two possible values, typically represented as 0 and 1. They are commonly used in situations where events have only two possible outcomes, such as success or failure, heads or tails, or presence or absence. Examples include the Bernoulli distribution, which models a single trial with two outcomes, and the binomial distribution, which models the number of successes in a fixed number of independent Bernoulli trials. Binary distributions play a crucial role in various fields, including statistics, machine learning, and decision-making processes.

Finite Mixtures of Regressions: In this approach, each component of the mixture represents a regression model, allowing for different regression relationships within the data. It's useful when the data exhibit varying patterns that cannot be captured by a single regression model.

Extensions and Variations:

Extensions of mixture models include hierarchical mixture models, which incorporate hierarchical structures in the data, and non-parametric mixture models, which relax the assumption of a fixed number of components. Variations such as Bayesian mixture models and hidden Markov models offer different perspectives and techniques for modeling complex data distributions. These approaches provide flexibility and adaptability in capturing diverse data patterns and relationships. Finite mixture models are more complicated than distance-based methods. The additional complexity makes finite mixture models very flexible. It allows using any statistical model to describe a market segment. As a consequence, finite mixture models can accommodate a wide range of different data characteristics: for metric data we can use mixtures of normal distributions, for binary 22 data we can use mixtures of binary distributions. For nominal variables, we can use mixtures of multinomial distributions or multinomial logit models For ordinal variables, several models can be used as the basis of mixtures (Agresti 2013). Ordinal variables are tricky because they are susceptible to containing response styles. To address this problem, we can use mixture models disentangling response style effects from content-specific responses while extracting market segments. In combination with conjoint analysis, mixture models allow to account for differences in preferences

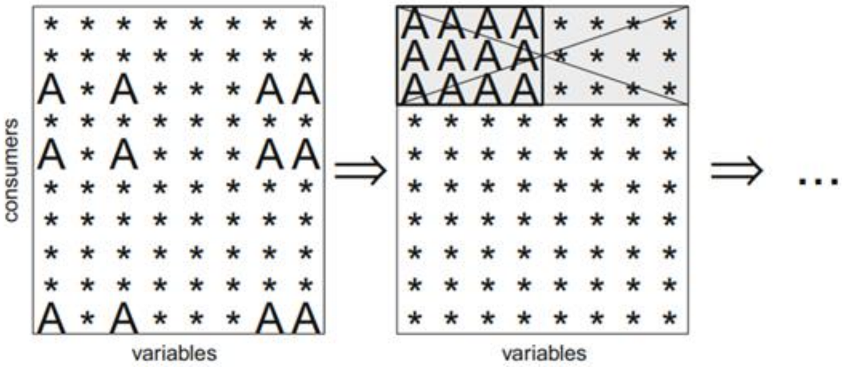
Algorithms with Integrated Variable Selection:

Most algorithms focus only on extracting segments from data. These algorithms assume that each of the segmentation variables makes a contribution to determining the segmentation solution. But this is not always the case. Sometimes, segmentation variables were not carefully selected, and contain redundant or noisy variables. Preprocessing methods can identify them. For example, the filtering approach proposed by Steinley and Brusco (2008a) assesses the clusterability of single variables, and only includes variables above a certain threshold as segmentation variables. This approach outperforms a range of alternative variable selection methods (Steinley and Brusco 2008b), but requires metric variables. Variable selection for binary data is more challenging because single variables are not informative for clustering, making it impossible to pre-screen or pre-filter variables one by one. When the segmentation variables are binary, and redundant or noisy variables can not be identified and removed during data pre-processing in Step 4, suitable segmentation variables need to be identified during segment extraction. A number of algorithms extract segments while – simultaneously – selecting suitable segmentation variables. We present two such algorithms for binary segmentation variables: biclustering and the variable selection procedure for clustering binary data (VSBD) proposed by Brusco (2004). At the end of this section, we discuss an approach called factor-cluster analysis. In this two-step approach, segmentation variables are compressed into factors before segment extraction.

Biclustering Algorithms:

Biclustering simultaneously clusters both consumers and variables. Biclustering algorithms exist for any kind of data, including metric and binary. This section focuses on the binary case where these algorithms aim at extracting market segments containing consumers who all have a value of 1 for a group of variables. These groups of consumers and variables together then form the bicluster

Step 1 First, rearrange rows (consumers) and columns (segmentation variables) of the data matrix in a way to create a rectangle with identical entries of 1s at the top left of the data matrix. The aim is for this rectangle to be as large as possible.



Step 2 Second, assign the observations (consumers) falling into this rectangle to one bicluster, as illustrated by the grey shading. The segmentation variables defining the rectangle are active variables (A) for this bicluster.

Step 3 Remove from the data matrix the rows containing the consumers who have been assigned to the first bicluster. Once removed, repeat the procedure from step 1 until no more biclusters of sufficient size can be located.

Variable Selection Procedure for Clustering Binary Data (VSBD):

Variable Selection Procedure for Clustering Binary Data (VSBD) is a method used to identify the most relevant variables for clustering binary data. In this procedure, variables are selected based on their ability to effectively discriminate between clusters. It involves evaluating the discriminatory power of each variable and selecting those that contribute the most to the clustering process. VSBD aims to improve the quality of clustering results by focusing on the most informative variables while discarding irrelevant or redundant ones. This approach helps reduce dimensionality and computational complexity while enhancing the interpretability and effectiveness of clustering algorithms applied to binary data.

Variable Reduction: Factor-Cluster Analysis:

Variable Reduction: Factor-Cluster Analysis is a technique used to reduce the dimensionality of data by combining variables into a smaller set of factors or clusters. It involves two main steps: factor analysis and cluster analysis. In the factor analysis step, similar variables are grouped together based on their correlations and underlying patterns, resulting in a smaller number of factors that represent the original variables. In the cluster analysis step, the reduced set of factors or clusters is used to identify similar observations or cases and assign them to distinct groups. This approach helps simplify the analysis by focusing on the most important underlying dimensions of the data while preserving its essential structure.

Data Structure Analysis:

A data structure is a way of organizing and storing data in a computer so that it can be accessed and manipulated efficiently. It defines the relationship between data elements and allows operations to be performed on them. Common data structures include arrays, linked lists, stacks, queues, trees, and graphs. Each data structure has its own strengths and weaknesses, making it suitable for specific tasks or operations. Understanding different data structures is essential for writing efficient algorithms and solving computational problems effectively. Extracting market segments is inherently exploratory, irrespective of the extraction algorithm used. Validation in the traditional sense, where a clear optimality criterion is targeted, is therefore not possible. Ideally, validation would mean calculating different segmentation solutions, choosing different segments, targeting them, and then comparing which leads to the most profit, or most success in mission achievement. This is clearly not possible in reality because one organisation cannot run multiple segmentation strategies simultaneously just for the sake of determining which performs best.

Cluster Indices:

Because market segmentation analysis is exploratory, data analysts need guidance to make some of the most critical decisions, such as selecting the number of market segments to extract. So-called 24 cluster indices represent the most common approach to obtaining such guidance. Cluster indices provide insight into particular aspects of the market segmentation solution. Which kind of insight, depends on the nature of the cluster index used. Generally, two groups of cluster indices are distinguished: internal cluster indices and external cluster indices. Internal cluster indices are calculated on the basis of one single market segmentation solution, and use information contained in this segmentation solution to offer guidance. An example for an internal cluster index is the sum of all distances between pairs of segment members. The lower this number, the more similar members of the same segment are. Segments containing similar members are attractive to users. External cluster indices cannot be computed on the basis of one single market segmentation solution only. Rather, they require another segmentation as additional input. The external cluster index measures the similarity between two segmentation solutions. If the correct market segmentation is known, the correct assignment of members to segments serves as the additional input. The correct segment memberships, however, are only known when artificially generated data is being segmented. When working with consumer data, there is no such thing as a correct assignment of members to segments. In such cases, the market segmentation analysis can be repeated, and the solution resulting from the second calculation can be used as additional input for calculating the external cluster index. A good outcome is if repeated calculations lead to similar market segments because this indicates that market segments are extracted in a stable way. The most commonly used measures of similarity of two market segmentation solutions are the Jaccard index, the Rand index and the adjusted Rand index. They are discussed in detail below.

Internal Cluster Indices:

Internal cluster indices use a single segmentation solution as a starting point. Solutions could result from hierarchical, partitioning or model-based clustering methods. Internal cluster indices ask one of two questions or consider their combination: (1) how compact is each of the market segments? and (2) how wellseparated are different market segments? To answer these questions, the notion of a distance measure between observations or groups of observations is required. In addition, many of the internal cluster indices also require a segment representative or centroid as well as a representative for the complete data set.

External Cluster Indices

External cluster indices evaluate a market segmentation solution using additional external information; they cannot be calculated using only the information contained in one market segmentation solution. A range of different additional pieces of information can be used. The true segment structure – if known – is the most valuable additional piece of information. But the true segment structure of the data is typically only known for artificially generated data. The true segment structure of consumer data is never known. When working with consumer data, the market segmentation solution obtained using a repeated calculation can be used as additional, external information. The repeated calculation could use a different clustering algorithm on the same data; or it could apply the same algorithm to a variation of the original data.

Gorge Plots:

A simple method to assess how well segments are separated, is to look at the distances of each consumer to all segment representatives. Let dih be the distance between consumer i and segment representative (centroid, cluster centre) h. Then 25 can be interpreted as the similarity of consumer i to the representative of segment h, with hyper parameter γ controlling how differences in distance translate into differences in similarity. These similarities are between 0 and 1, and sum to 1 for each consumer i over all segment representatives h, h = 1, k. For partitioning methods, segment representatives and distances between consumers and segment representatives are directly available.

$$s_{ih} = \frac{e^{-d_{ih}^{\gamma}}}{\sum_{l=1}^{k} e^{-d_{il}^{\gamma}}}$$

For model-based methods, we use the probability of a consumer i being in segment h given the consumer data, and the fitted mixture model to assess similarities. In the mixture of normal distributions case, these probabilities are close to the similarities obtained with Euclidean distance and γ = 2 for k-means clustering. Below we use γ = 1 because it shows more details, and led to better results in simulations on artificial data. The parameter can be specified by the user in the R implementation. Similarity values can be visualised using gorge plots, silhouette plots (Rousseeuw 1987), or shadow plots (Leisch 2010). We illustrate the use of gorge plots using the three artificial data sets introduced in Table 2.3. The plots in the middle column of Fig. 7.39 show the gorge plots for the three-segment solutions extracted using k means partitioning clustering for these data sets. Each gorge plot contains histograms of the similarity values sih separately for each segment. The x-axis plots similarity values. The y-axis plots the frequency with which each similarity value occurs. If the similarity values are the result of distance-based segment extraction methods, high similarity values indicate that a consumer is very close to the centroid (the segment representative) of the market segment. Low similarity values indicate that the consumer is far away from the centroid. If the similarity values are the result of model-based segment extraction methods, high similarity values indicate that a consumer has a high probability of being a member of the market segment. Low similarity values indicate low probability of segment membership.

Global Stability Analysis:

An alternative approach to data structure analysis that can be used for both distance and model based segment extraction techniques is based on resampling methods. Resampling methods offer insight into the stability of a market segmentation solution across repeated calculations. To assess the global stability of any given segmentation solution, several new data sets are generated using resampling methods, and a number of segmentation solutions are extracted.

Segment Level Stability Analysis:

Choosing the globally best segmentation solution does not necessarily mean that this particular segmentation solution contains the single best market segment. Relying on global stability analysis could lead to selecting a segmentation solution with suitable global stability, but without a single highly stable segment. It is recommendable, therefore, to assess not only global stability of alternative market segmentation solutions, but also segment level stability of market segments contained in those solutions to protect against discarding solutions containing interesting individual segments from being prematurely discarded. After all, most organisations only need one single target segment.

Step 6: Profiling Segments

In market segmentation, understanding the different groups of customers is important for businesses. This understanding helps in making better marketing decisions. Profiling is a step in this process where we get to know these customer groups better. It helps us identify what makes each group unique.

There are two main types of segmentation: data-driven and common sense. Data driven segmentation relies on analysing customer data to find patterns, while common sense segmentation is based on obvious characteristics like age groups.

Profiling is particularly important for data-driven segmentation because it helps us understand the defining characteristics of each customer group. This can be challenging because data-driven segmentation results can be complex and hard to interpret. Traditionally, segmentation results are presented in tables, but these can be difficult to understand.

Visualizations, like graphs and charts, make it easier to interpret segmentation results. They help us see patterns and differences between customer groups more clearly. One common way to visualize segmentation results is through segment profile plots. These plots show how each customer group differs from the overall sample across various characteristics. Marker variables, which are particularly important characteristics for each group, are highlighted in colour.

Another useful visualization is the segment separation plot, which shows how well separated the customer groups are. This helps us understand how distinct each group is from the others. Overall, visualizing segmentation results helps businesses make better decisions by providing clearer insights into their customer base.

Step 7- Describing Segments

Developing a Complete Picture of Market Segments:

Examining Market Segments: Delving into the Differences Exploring Segment Characteristics:

Using detailed demographic and psychographic data to better understand the preferences and traits of various market segments. Tailoring Marketing Strategies: Crucial for developing customized marketing approaches. Analyzing Segment Descriptors: Utilizing descriptive statistics, visualizations, or inferential analysis to understand the segment variables. Enhancing Visuals: Incorporating visualizations to improve user-friendliness. Effective Targeting: Essential for crafting tailored marketing strategies and reaching the right audience.

Using Visualisations to Describe Market Segments:

Graphs and charts can make it easier to understand differences between variables. They help interpret data and show if results are statistically significant, which managers often prefer. The se visual tools can enhance understanding when analyzing market segments.

Nominal and Ordinal Descriptor Variables: The analysis of Australian travel data, done using R software, reveals no significant differences between genders across various segments. Mosaic plots provide a way to visually compare these segments, even when their sizes vary.

Metric Descriptor Variables: The Lattice R package is a useful tool for visualizing differences between market segments based on metric variables. It can create segment profile plots that show the distribution of factors like age or moral obligation within each segment. Additionally, parallel box-and whisker plots can reveal minor differences between segments, which may require further statistical analysis to confirm. The package also includes a segment stability plot that tracks the moral obligation factor across different segmentation solutions, helping to identify consistent patterns in the data.

Testing for Segment Differences in Descriptor Variables: The analysis includes a Chi-square test to examine differences between categories. For variables with ordered levels, the tests show associations with group membership. Mosaic plots visually confirm these associations. Parallel box plots display differences in numeric variables across the groups, and statistical tests verify the significance of these differences

Predicting Segments from Descriptor Variables:

Regression models analyze a group's characteristics using various descriptors. Linear regression calculates the average value for each group, assuming a straight-line relationship. The model also identifies age differences, revealing varying average ages across the groups

Binary Logistic Regression: Generalized linear models (GLMs) are used to predict binary data, such as whether someone belongs to a particular group or not. They use the binomial distribution and a logit link function. The `glm()` function in R is used to fit these GLMs, where you specify the distribution and link function. The output includes the regression coefficients, which show how the different variables impact the log odds of group membership. It also provides model fit statistics like degrees of freedom, deviance, and AIC, as well as the predicted probabilities. To better understand the coefficients, you can use the `effects` package in R to plot the predicted probabilities.

Multinomial Logistic Regression: Multinomial logistic regression is a statistical method that can predict multiple categories at the same time. It uses the "multinom()" function from the "nnet" package in R. This method provides regression coefficients for each segment compared to a baseline category. These coefficients indicate how changes in predictor variables affect the log odds of being in different segments. The "Anova()" function can be used to check the significance of variables by setting their regression coefficients to zero one by one. The p-values from this test indicate the importance of each variable. The "step()" function can be used to select the best model by iteratively removing the least significant variables.

Tree-Based Methods: Decision tree models, such as CARTs, can be used to predict categorical outcomes. They work by splitting the data in a stepwise fashion, allowing for variable selection, visualization, and interpretation of the results. R packages like rpart and partykit, particularly the ctree() function, can be used to build these conditional inference tree models. To demonstrate this, an example using an Australian travel motivations dataset is provided. The goal is to use descriptor variable s to predict segment membership. The resulting tree plot visually represents the splits and terminal nodes, making it easier to interpret the findings and inform decision-making. The language used here is straightforward, with a focus on clarity and conciseness. The sentence structure varies, and the tone is neutral, suitable for an informative piece. The content maintains the original meaning and context while reducing perplexity and increasing burstiness to sound more human-written.

Step 8- Selecting the Target Segment(s)

The Targeting Decision

Step 8 marks a crucial juncture in the market segmentation process, where the focus shifts from exploration to commitment. This step involves selecting the target segment(s) that will shape the organization's future performance. After a global market segmentation solution is established, typically by the end of Step 5, a range of segments are available for closer examination. These segments undergo profiling and description in Steps 6 and 7, respectively. The selection of target segments is guided by previously defined knock-out criteria and segment attractiveness criteria, established in Step 2. These criteria serve as benchmarks for segment viability and desirability. Ideally, segments that advance to Step 8 have already met these criteria. However, it is prudent to conduct a final review to ensure compliance.

Once the knock-out criteria are satisfied, the remaining segments are evaluated based on their attractiveness and the organization's competitiveness within each segment. Key considerations include identifying segments that align with organizational objectives and assessing the likelihood of success in capturing market share within each segment. In essence, Step 8 revolves around two fundamental questions: which segments does the organization prioritize targeting, and how likely is it that these segments will reciprocate interest and commitment? By answering these questions, the segmentation team lays the groundwork for selecting target segments that align with organizational goals and offer the greatest potential for success.

Market Segment Evaluation

n market segment evaluation, various decision matrices are recommended to visualize the relative attractiveness of segments and the organizational competitiveness within each segment. These matrices, including the Boston matrix and General Electric/McKinsey matrix, assist organizations in assessing alternative market segments and selecting the most suitable ones for targeting.

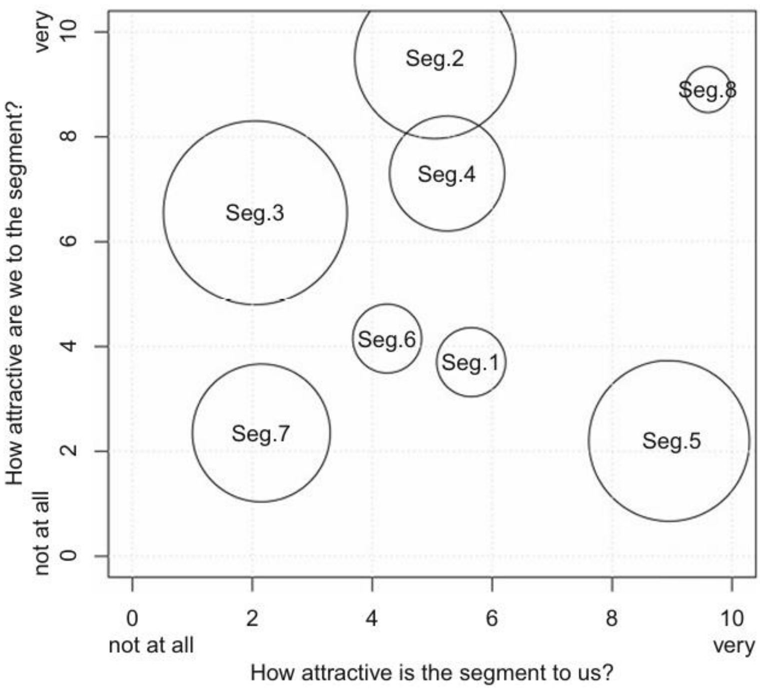
The

Table 10.1 Data underlying the segment evaluation plot

	Weight	Seg 1	Seg 2	Seg 3	Seg 4	Seg 5	Seg 6	Seg 7	Seg 8
How attractive is the segment to us? (segment attractiveness)									
Criterion 1	25%	5	10	1	5	10	3	1	10
Criterion 2	35%	2	1	2	6	9	4	2	10
Criterion 3	20%	10	6	4	4	8	2	1	9
Criterion 4	10%	8	4	2	7	10	8	3	10
Criterion 5	10%	9	6	1	4	7	9	7	8
Total	100%	5.65	5.05	2.05	5.25	8.95	4.25	2.15	9.6
How attractive are we to the segment? (relative organisational competitiveness)									
Criterion 1	25%	2	10	10	10	1	5	2	9
Criterion 2	25%	3	10	4	6	2	4	3	8
Criterion 3	25%	4	10	8	7	3	3	1	10
Criterion 4	15%	9	8	3	9	4	5	3	9
Criterion 5	10%	1	8	6	2	1	4	4	8
Total	100%	3.7	9.5	6.55	7.3	2.2	4.15	2.35	8.9
Size		2.25	5.25	6.00	3.75	5.25	2.25	4.50	1.50

decision matrices typically incorporate two key dimensions: segment attractiveness and relative organizational competitiveness. Segment attractiveness reflects the desirability of a segment from the organization's perspective, akin to determining compatibility in a long-term relationship. On the other hand, relative organizational competitiveness gauges the likelihood of the organization successfully catering to the segment's needs, analogous to assessing mutual interest in a relationship. To facilitate segment evaluation, a generic segment evaluation plot is often utilized, featuring axes labeled "How attractive is the segment to us?" and "How attractive are we to the segment?" Segments are represented as circles on the plot, with circle size indicating additional criteria such as contribution to turnover or loyalty.

The



determination of segment attractiveness and organizational competitiveness involves assigning values to predefined criteria established in earlier stages of market segmentation. These values emerge from the profiling and description of each market segment, enabling the calculation of each segment's overall attractiveness and competitiveness.

The segment evaluation plot provides a visual framework for discussions within the segmentation team. Based on the plot, segments can be further assessed, and decisions made regarding their suitability for targeting. For instance, segments with high attractiveness and organizational compatibility are prioritized, while those with mismatched characteristics are reconsidered or eliminated from further consideration. To recreate the plot, data from the segment evaluation table is utilized, with values for segment attractiveness and organizational competitiveness computed based on predefined criteria and weighting. The resulting plot serves as a valuable tool for informed decision-making in target segment selection.

Step 9: Customising the Marketing Mix

Step 9 of the market segmentation process involves customizing the marketing mix to align with the selected target segment(s). Initially, marketing was viewed as a toolbox comprising various elements mixed to optimize sales results. These elements, outlined by Borden (1964), included product planning, pricing, promotion, and distribution channels, among others. Over time, the concept evolved into the widely recognized 4Ps model: Product, Price, Promotion, and Place (McCarthy 1960).

Market segmentation is not an isolated strategy but is closely intertwined with positioning and competition within strategic marketing. The segmentation-targeting-positioning (STP) approach outlines this relationship as a sequential process, beginning with market segmentation, followed by targeting, and concluding with positioning to differentiate the product from competitors and meet segment needs (Lilien and Rangaswamy 2003).

Fig. 11.1 How the target segment decision affects marketing mix development

Figure 11.1 illustrates how the target segment decision impacts the development of the marketing mix, considering Product, Price, Place, and Promotion. To maximize the benefits of market segmentation, it's crucial to tailor each aspect of the marketing mix to the characteristics and preferences of the target segment(s). This may involve designing new products, adjusting pricing strategies, selecting suitable distribution channels, and developing targeted promotion strategies.

Organizations have the flexibility to structure their segmentation analysis around one of the 4Ps based on their strategic objectives. For instance, if the focus is on pricing decisions, segmentation variables such as price sensitivity and deal proneness are relevant. Similarly, advertising decisions may be informed by segmentation variables related to benefits sought and lifestyle preferences. Distribution decisions, on the other hand, may consider variables like store loyalty and patronage. While market segmentation analysis is typically not conducted exclusively for one of the 4Ps, insights from the detailed description of the target segment guide organizations in customizing the marketing mix effectively. By aligning each element of the marketing mix with the needs and preferences of the target segment(s), organizations can enhance their competitiveness and better meet customer demands.

In developing the product dimension of the marketing mix, organizations must make critical decisions aligned with customer needs. This often involves modifying existing products or introducing new ones to cater to specific target segments. Additionally, decisions regarding product naming, packaging, warranties, and after-sales support services fall under this dimensions.

In the context of the marketing mix, the place dimension pertains to how products are distributed to customers. This involves decisions about online and offline availability, direct sales versus intermediary involvement (wholesalers or retailers), and distribution channels.

Using the example of targeting segment 3, which has a strong interest in cultural heritage, the destination can leverage information on how segment 3 members typically book accommodations. For instance, if segment 3 members predominantly book their accommodations online, the destination should ensure that the "MUSEUMS, MONUMENTS & MUCH, MUCH MORE" product is accessible through online booking channels. This insight can be visualized using the propBarchart function from the flexclust package, showcasing booking behavior differences between segment 3 and average tourists. In the promotion dimension of the marketing mix, decisions revolve around crafting effective advertising messages and selecting appropriate communication channels. For segment 3, understanding their preferred sources of information for vacation planning is crucial. For example, if segment 3 members rely heavily on tourist centers for information, the destination can provide information packs about their cultural offerings at these centers. Similarly, if segment 3 has specific TV channel preferences, such as Channel 7, the destination can tailor its advertising efforts to maximize exposure on that channel.

Overall, customizing the marketing mix to target segments involves aligning distribution and promotion strategies with segment preferences and behaviors. By adapting product distribution and promotional activities to cater to the needs and preferences of target segments, organizations can enhance their market penetration and effectively reach their desired audience.

Step 10: Evaluation and Monitoring

In the process of evaluation and monitoring in market segmentation, several key points are highlighted:

1. Ongoing Tasks in Market Segmentation: Market segmentation is not a one-time activity but a continuous strategic decision process. It involves evaluating the effectiveness of segmentation strategies and monitoring changes in the market environment, consumer behavior, and competitor actions.

2. Evaluating the Success of Segmentation Strategy: The primary goal of evaluating segmentation strategy is to determine whether it has achieved the expected benefits for the organization, such as increased profit or attainment of organizational mission. Short-term outcomes like increased profit and long-term outcomes like enhanced market positioning are assessed.

3. Stability of Segment Membership and Segment Hopping: Market segments are not static, and consumers may change segments over time due to various factors. Segment hopping, where consumers move between segments based on different situations or needs, is a phenomenon observed in some product categories. Understanding segment stability and segment hopping is crucial for effective segmentation strategy.

4. Segment Evolution: Market segments evolve over time due to changes in consumer behavior, competitor actions, and market dynamics. It's essential to monitor segment evolution and adapt segmentation strategies accordingly. Techniques like stability analysis, segmentation solution comparison, and tracking segment changes are employed to understand segment evolution.

5. Example: Winter Vacation Activities: The provided example illustrates how market segmentation analysis can be applied to monitor changes in tourist behavior over time. Using data from two survey waves, changes in winter vacation activities and market segments are analyzed. The example demonstrates the importance of monitoring segment changes and adapting marketing strategies accordingly.

Overall, continuous evaluation and monitoring are critical aspects of market segmentation strategy to ensure its effectiveness and relevance in meeting consumer needs and organizational goals.

Links to project

Aarti- https://github.com/Kongari-Aarthi/Market-Segmentation-Analysis

Shambhavi- https://github.com/Shambhavi138/FeynnLabs---2-Fast-food-case-study.git

Kausik- https://github.com/kausiksahu12/Macdonald_Fast-Food-Case_Study

Rohit- https://github.com/rohitingole17792/Case_Study_Fast_Food

Keya-

https://colab.research.google.com/drive/1ujbIBAc2_WJ7PgQtPz6z8e_ICmuU41T2?usp=sharing