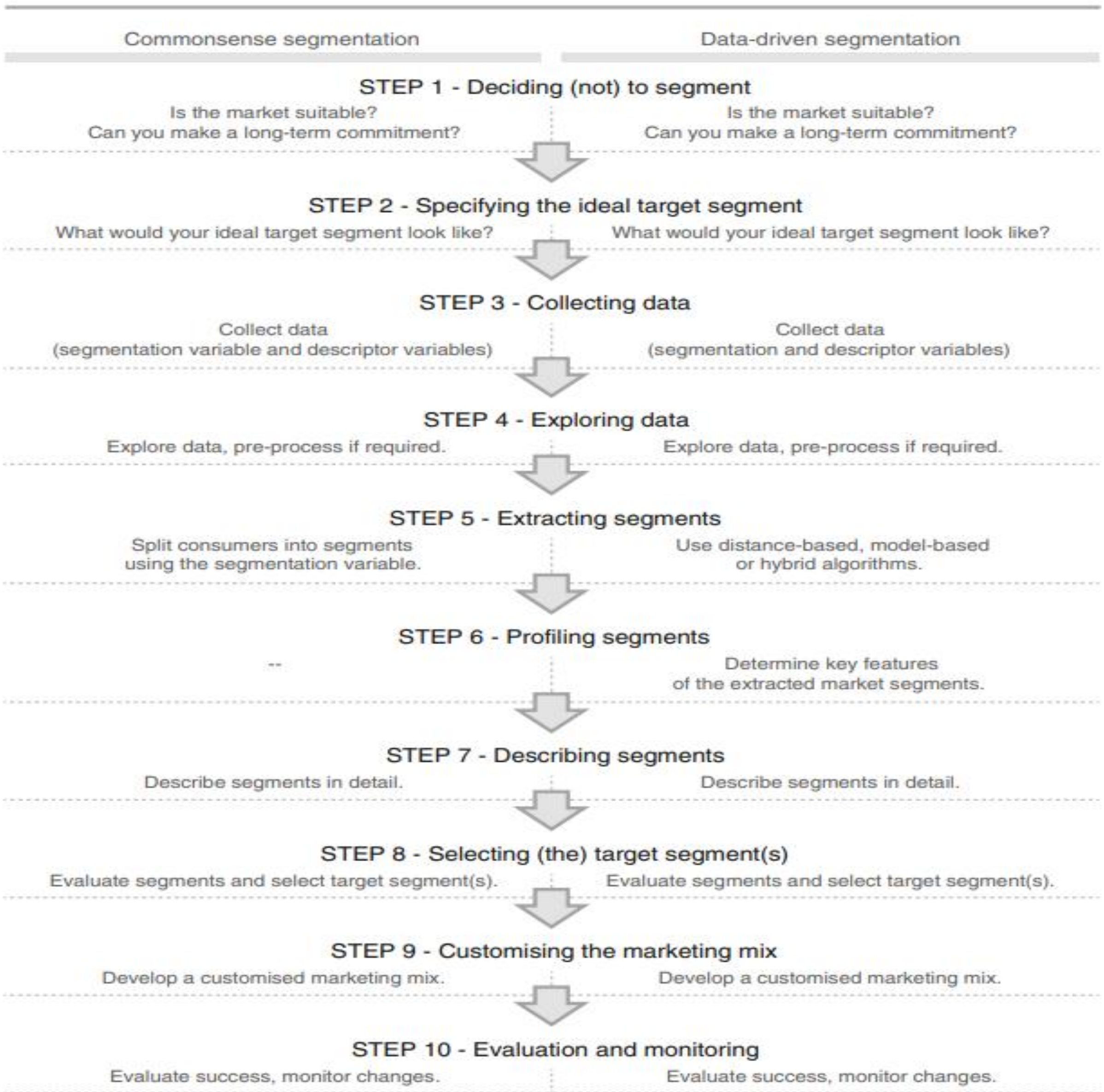
MARKET SEGMENTATION – CASE STUDY

Introduction:

Market segmentation is a crucial tool for marketing managers as it helps them identify a target market for their product, and create a marketing plan that caters to that market. Segmentation is a critical aspect of strategic marketing and is instrumental in achieving success. Successful businesses base their strategies on segmentation, and it plays a central role in marketing decisions. Market segmentation helps organizations identify where their revenue is coming from and where it will be in the future. It is also helpful in understanding the organization's strengths compared to its competitors. However, it is a costly process that requires significant human and financial resources.

Steps of Market Segmentation:



Step 1: Deciding (not) to Segment

Marketing segmentation is an important aspect of a marketing strategy for many organizations. However, before deciding to invest in market segmentation analysis, it is crucial to understand the implications of doing so. One key implication is that the organization must commit to the marketing strategy for a long time, as implementing segmentation requires substantial changes that can be costly. According to Cahill, market segmentation is not free as it involves expenses such as research, fielding surveys, focus groups, designing multiple packages, advertisements, and communication messages. Therefore, it is advisable to pursue segmentation only if the increased sales are sufficient to generate more profit than the marketing costs.

Marketing segmentation can result in the development of new products, modifications to existing products, changes in pricing and distribution channels, and alterations to communication strategies. These changes within an organization can take time to adjust to and can also result in additional costs, such as targeting more consumer segments. It is important to note that these changes are not related to the product itself, but rather to marketing efforts. Therefore, it is recommended that the potential of market segmentation strategy be investigated at the highest executive level as it requires a long-term commitment and should be communicated throughout the entire organization.

Implementation Barrier:

- The primary barrier that hinders the success of an organization’s market segmentation is the lack of leadership qualities in the senior management. Their improper proactive approach towards commitment and involvement in providing proper resources to obtain meaningful insights can negatively impact the company's success. Therefore, senior management needs to exhibit strong leadership qualities and provide adequate resources to achieve successful market segmentation.
- The success of an organization's market segmentation is hindered by a lack of market and consumer knowledge, creative ideas, and a poor office environment due to communication issues, lack of information sharing, and office politics.
- The team is facing challenges with market segmentation due to a lack of training provided by senior management. They fail to understand the basics of market segmentation and its consequences, which makes it difficult for them to develop effective marketing strategies.
- Lack of competent marketing professionals can affect the success of market segmentation and marketing strategies.
- Another barrier to effective market segmentation is the financial barrier. Many organizations do not have sufficient financial resources to implement market segmentation properly, as several aspects of market segmentation require additional human and other resources.

Step 2: Specifying the Ideal Target Segment

- **Segment Evaluation Criteria:**

The third layer of Market Segmentation Analysis heavily relies on user input. Therefore, the user is not only involved in the beginning stages of Market Segmentation or the development of market mix strategy but is also involved in the entire process. The user's technical expertise is crucial to the success of Marketing Segmentation, and they are required to possess the best technical aspects throughout the process.

In Step 2, the organization should rely on two sets of criteria. The first is the Knockout Criteria, which includes non-negotiable segments of the market that are essential. The second is the Attractiveness Criteria, which evaluates the relative attractiveness of the remaining market segments.

  - Knock Out Criteria:
    - It should be Homogenous means members should be Similar.
    - Distinct: Members Should be Different from Other Segments.
    - Segments Should have a large number of members so that it’s considerable to spend money for customizing market strategy for them.
    - Satisfying the segment's Member's Needs.
    - Members should be easily identified in the marketplace.
    - Members of Segments should be accessible so that a customised market mix should benefit them.
  - Attractiveness Criteria: These criteria are not binary; attractiveness determines the target based on market segmentation.

- **Implementing a Structured Process:**

The criteria for market segmentation should be finalized only after a detailed investigation by the organization as there is no set standard to set criteria. The attractiveness and competitiveness of a segment are determined by the segmentation working team. The team responsible for market segmentation, along with the initial team, should propose a solution and report to the senior management members from all units of the organization for discussion. This is important as each unit has different perspectives and is a stakeholder of the organization. Therefore, the final marketing segmentation criteria should be modified and advised by all units before implementation. Proper information and data collection should be performed as per the required criteria to fulfil the needs of each unit.

Step 3: Collecting Data

- **Segmentation Variable:**

Empirical data is used in market segmentation to identify or construct market segments and describe them in detail. In commonsense segmentation, only one characteristic is used to divide the sample into segments. Other personal characteristics act as descriptive variables, aiding in the development of targeted marketing strategies. In contrast, data-driven market segmentation employs multiple variables to identify naturally occurring or artificially constructed segments that are beneficial to the organization's objectives.

Market segmentation is the process of dividing a market into groups of consumers who share similar preferences. In traditional segmentation, marketers use factors like gender, age, and income, whereas in data-driven segmentation, marketers use preferences to divide the market into distinct segments.

To achieve meaningful market segmentation, businesses need high-quality empirical data. This data is vital for correctly assigning individuals to segments and informing tailored product offerings, pricing strategies, distribution channels, and advertising avenues.

Data for segmentation studies can be sourced from surveys, observations like scanner data, or experimental studies. It's essential to explore various data sources, prioritizing those that best capture actual consumer behavior for optimal segmentation outcomes.

- **Segmentation Criteria:**

Segmentation is a critical marketing decision. Common criteria include geographic, socio-demographic, psychographic, and behavioral factors. Choosing the best criterion can be challenging without clear guidelines. Opt for the simplest approach possible. Use demographic segmentation if it fits the product or service. What matters most is what works effectively for the product or service at a minimal cost.

  - **Geographic Segmentation:**

    Geographic segmentation focuses on identifying consumers based on their location of residence, facilitating targeted communication through local media channels. While it has limitations, it has proven effective in various contexts. However, geographical proximity does not necessarily imply shared consumer preferences or characteristics. Despite its challenges, geographic segmentation has seen a resurgence in international market studies. Haverila's (2013) study on mobile phone users across national borders exemplifies such international segmentation endeavors.

  - **Socio-Demographic Segmentation:**

    Socio-demographic segmentation is widely used in various industries to identify consumer groups based on factors like age, gender, income, and education. However, socio-demographic criteria may not always be the most significant determinant of consumer behavior. Research suggests that values, tastes, and preferences hold more significance in influencing buying decisions than socio-demographic factors. Therefore, a deeper understanding of consumer attitudes and motivations is important in segmentation strategies.

- **Psychographic Segmentation:**

  Psychological dimensions accurately. Psychographic segmentation categorizes people based on psychological criteria such as beliefs, interests, preferences, aspirations, or desired product benefits. It's more complex than geographic or socio-demographic factors, but provides deeper insights into consumer behavior. Psychographic segmentation studies use multiple variables to capture psychological factors effectively. Its effectiveness depends on the reliability and validity of the measures used to capture psychological factors.

- **Behavioural Segmentation:**

  Behavioral segmentation groups consumers based on their observed or reported actions. Studies show that it's more effective than using geographical factors. It focuses on actual consumer behavior, providing valuable insights for marketing. It eliminates the need to develop measures for psychological constructs. However, obtaining behavioral data can be challenging. Despite this, it offers valuable insights for segmentation analysis and marketing strategy development.

- **Data from Survey Studies:**

  Market segmentation analyses rely on survey data due to its affordability and ease of collection. However, survey data is susceptible to various biases, potentially compromising the quality of segmentation solutions derived from the analysis. Therefore, organizations must be mindful of these biases and take steps to mitigate their impact to ensure accurate and reliable results.

  - **Choice of Variables:**

    Carefully selecting variables for segmentation is crucial for quality market segmentation. In data-driven segmentation, including only relevant variables is important to avoid respondent fatigue and to ensure that algorithms can identify the correct segmentation solution. Noisy or masking variables can obstruct algorithms and should be avoided. Redundant questions can impede segmentation analysis. Qualitative research can provide valuable insights into respondents' beliefs, enhancing the robustness of segmentation analysis.

- **Response Options:**

  The options provided to survey respondents can impact segmentation analysis. Binary or metric data are ideal for analysis. Nominal and ordinal data pose challenges. Binary or metric response options should be provided. Visual analogue scales are a good option. Research suggests that binary response options outperform ordinal ones, and provide valuable insights without sacrificing analytical clarity.

- **Response Styles:**

  Survey data collection can be biased due to response biases and styles, which can distort segmentation outcomes. It's important to minimize the influence of response styles during data collection to ensure accurate segmentation. Additional analyses may be required to exclude segments affected by response biases, or respondents exhibiting consistent response styles may be filtered out.

- **Sample size:**

  Market segmentation analysis requires an adequate sample size, but there are no specific recommendations for sample size. Studies suggest a sample size of at least 2p or 10 times the number of segmentation variables times the number of segments. Dolnicar et al. (2014, 2016) recommend a sample size of at least 60p for basic scenarios and 70p for more complex ones. Increasing sample size can improve segmentation accuracy, but the degree of improvement varies depending on market and data characteristics. It is crucial to have a large, high-quality sample with unbiased data for accurate market segmentation analysis. A guideline is to have at least 100 respondents for each segmentation variable, with careful attention to data quality and characteristics.

- **Data from Internal Sources:**

  Organizations use internal data sources, such as scanner data from grocery stores, booking data from airline loyalty programs, and online purchase data, for market segmentation analysis. These data reflect actual consumer behavior and are easily available. However, internal data can be biased, lacking information about potential future customers. To ensure unbiased analysis, organizations should complement internal data with external sources.

- **Data from Experimental Studies:**

  Experimental data from field or laboratory experiments is a valuable source for market segmentation analysis. It can involve testing consumer responses to ads or conducting choice experiments and conjoint analyses. These data provide insights into consumer behavior and preferences, offering potential segmentation criteria for market analyses.

Step 4: Exploring Data

- **Take a quick look at the data:**

  After data collection, the first thing is we need to explore the data in order to understand the overall structure of the data so that we get suitable idea on what data are we dealing with for market segmentation, if we should clean and pre-process the data and finally it helps to select a suitable algorithm for market segment extraction. Exploratory data analysis cleans and pre-processes collected data and helps identify measurement levels, segmentation variable distributions and dependency structures of segmentation variables. Pre-processing and preparation of data may be necessary for different segmentation algorithms and exploration results provide insights into their suitability.

- **Data Cleaning:**

  After a quick glance at the data the next thing is to check errors in the data and clean it. Bad data results in poor market segments. So we need to check for the errors and clean the data before analyzing it. It involves checking if all values have been recorded correctly and has consistent labels for categorical variables. For numeric variables, the range of reasonable values is known in advance, making it easy to identify any outliers or data errors. Only reasonable values should appear in the data, and any other values need to be corrected during the cleaning process. Data cleaning involves checking for wrong format, removing incorrect and inaccurate values and removing duplicate values. Data cleaning ensures a better result in market segment extraction.

- **Descriptive Analysis**:

  Once the data is cleaned, we need to analyze the data to understand it. Understanding data is crucial for accurate interpretation of market segments. We can analyze the data using descriptive statistics and using graphic representations. In descriptive statistics we calculate min, max, range, quartiles, standard deviation and mean for numeric variables in order to find reasonable values and remove outliers if any. It also helps to know distribution of the data and to treat skewness. Graphical methods are also used to understand the segmentation variables. For numeric data we can use histograms to understand distribution of the data, box plots to identify outliers and scatter plots to find relationship between the segmentation variables. For categorical variable, Bar plots is used to get frequency counts, while mosaic plots illustrate the association of multiple categorical variables.

- **Pre-processing:**

  Next step is to pre-process the data making it suitable for algorithms. Pre-processing and preparation of data is crucial in market segment analysis. For categorical variable, two pre-processing procedures are often used. One is merging levels of categorical variables before further analysis, the other one is converting categorical variables to numeric ones by one hot encoding or label encoding. For numerical variable, different range of values of a segmentation variable affects various algorithms and makes it difficult to extract segments. So we need to normalize the data for better results.

- **Principal Components Analysis:**

  It is a pre processing step where we treat higher dimensionality and multicollinearity problems. If the data has too many segmentation variables and are correlated with each other we can use pca to treat it. If data has more segmentation variables it makes the algorithm to find optimal results in higher dimension so it is reasonable to use pca in that case.

Step 5: Extracting Segments

Market segmentation is the process of dividing a heterogeneous market into smaller, more homogeneous segments based on certain characteristics or criteria. These segments typically share similar needs, preferences, or behaviors, allowing companies to tailor their marketing strategies and offerings to better meet the specific requirements of each segment. By understanding the distinct characteristics of different segments, businesses can effectively target their products or services, optimize their marketing efforts, and enhance customer satisfaction and loyalty.

Market segmentation helps companies identify and prioritize target markets, design customized marketing campaigns, allocate resources efficiently, and differentiate themselves from competitors. It is a fundamental concept in marketing strategy and is often part of a broader strategic approach that includes targeting and positioning to maximize the effectiveness of marketing initiatives.

Overall, market segmentation enables businesses to better understand their customers, improve their competitiveness, and ultimately drive growth and profitability in a dynamic and diverse marketplace.

- **Grouping Consumers**

Data-driven market segmentation analysis is exploratory in nature due to the unstructured consumer data sets. Consumer preferences are spread across the entire plot rather than forming clear groups, making the results of segmentation methods highly dependent on the assumptions made by the extraction algorithm and the underlying data. Segmentation methods shape the segmentation solution, and various clustering methods are commonly used to extract market segments. The choice of clustering method should match the data analytic features with the desired requirements, as different algorithms impose structures on the extracted segments.

For instance, k-means cluster analysis aims to find compact clusters covering a similar range in all dimensions, which may overlook existing patterns like spirals in the data. On the other hand, single linkage hierarchical clustering constructs snake-shaped clusters, making it more suitable for identifying complex structures like spirals. However, the effectiveness of an algorithm depends on the data structure, and there is no one-size-fits-all solution. Understanding the interaction between data and algorithm is crucial for extracting meaningful market segments.

The chapter provides an overview of popular extraction methods in market segmentation, categorizing them into distance-based and model-based methods. It emphasizes the importance of exploring and comparing alternative segmentation solutions to arrive at a suitable final solution. Factors such as data set characteristics, segment similarities and differences, and the number of segments influence algorithm selection. Additionally, the scale level of segmentation variables and the characteristics consumers should have in common play a significant role in determining the most appropriate extraction algorithm.

- **Distance Based Methods**

Distance-based methods are a category of techniques commonly used in market segmentation to measure the similarity or dissimilarity between individuals or objects based on their characteristics or attributes. These methods play a crucial role in clustering consumers into segments with similar needs, preferences, or behaviors. In the context of market segmentation for tourists with similar activity patterns during vacations, distance-based methods help quantify the differences in vacation activity preferences among individuals.

  - **Distance Measures:**

    The section provides an overview of distance measures commonly used in cluster analysis and market segmentation. It explains that distance measures are functions with two arguments (two vectors) that calculate the distance between them. The criteria for a distance measure include symmetry, where the distance between x and y is the same as between y and x, and the distance of a vector to itself is 0. The triangle inequality states that the distance from x to z via an intermediate point y is less than or equal to the direct distance from x to z.

    The section then introduces three common distance measures used in market segmentation analysis:

    - **Euclidean distance:** Represents the straight-line distance between two points in space, using the square root of the sum of squared differences between corresponding elements of the vectors.
    - **Manhattan distance:** Represents the distance between two points on a grid, calculated as the sum of absolute differences between corresponding elements of the vectors.
    - **Asymmetric binary distance:** Specifically for binary vectors, it calculates the proportion of common 1s over all dimensions where at least one vector contains a 1.

The section emphasizes that Euclidean distance is the most commonly used measure in market segmentation analysis, while Manhattan distance considers grid-like paths. Asymmetric binary distance focuses on the proportion of common 1s in binary vectors. The section also mentions the R function dist() for calculating distances, defaulting to Euclidean distance if not specified.

- **Hierarchical Methods**

Hierarchical methods refer to clustering techniques that organize data points into a hierarchy of clusters. These methods create a tree-like structure, known as a dendrogram, where the leaves represent individual data points, and the root represents a single cluster containing all data points. Hierarchical methods can be broadly categorized into two main types: agglomerative and divisive.

  - **Agglomerative Hierarchical Methods:**

    Agglomerative methods start with each data point as a separate cluster and iteratively merge the closest clusters based on a chosen distance metric until all data points belong to a single cluster. The process continues until a stopping criterion is met, such as a predetermined number of clusters or a specific level of dissimilarity. Common linkage criteria used to determine the distance between clusters include:

    - Single linkage: Based on the minimum distance between points in different clusters.
    - Complete linkage: Based on the maximum distance between points in different clusters.
    - Average linkage: Based on the average distance between points in different clusters.
    - Ward's method: Minimizes the total within-cluster variance when merging clusters.

  - **Divisive Hierarchical Methods:**

    Divisive methods take the opposite approach by starting with all data points in a single cluster and recursively splitting clusters into smaller clusters until each data point is in its own cluster. This top-down approach requires determining the optimal point to split clusters, often based on a chosen criterion like maximizing within-cluster homogeneity or minimizing within-cluster variance.

Hierarchical methods offer several advantages in market segmentation analysis:

- They provide a visual representation of the clustering process through dendrograms, aiding in the interpretation of cluster relationships.
- The hierarchical structure allows for the exploration of different numbers of clusters without the need to pre-specify the exact number.
- They can capture nested structures in the data, where smaller clusters are subsets of larger clusters.

However, hierarchical methods can be computationally intensive, especially with large datasets, and may not be as scalable as other clustering algorithms like K-means. Additionally, the choice of linkage criteria and distance metrics can significantly impact the clustering results, requiring careful consideration based on the characteristics of the data.

- **Partitioning Based Method**

Hierarchical clustering methods are effective for small datasets with up to a few hundred observations due to their ability to create a nested sequence of partitions. However, for larger datasets with more than 1000 observations, creating a single partition using clustering methods is more practical. In such cases, partitioning based methods are used. For instance, in a dataset with 1000 consumers, hierarchical clustering would need to compute 499,500 distances for the pairwise distance matrix. On the other hand, a partitioning clustering algorithm aiming for five market segments would only need to calculate between 5 and 5000 distances at each step. When extracting only a few segments, it is more efficient to optimize the algorithm for that specific goal rather than constructing the entire dendrogram and then cutting it into segments heuristically.

  - **k-Means and k-Centroid Clustering**

    The k-means clustering algorithm is a popular partitioning method used for market segmentation. It divides consumers into subsets (market segments) based on their similarities, with the representative of each segment known as the centroid. The algorithm involves five steps: specifying the number of segments, randomly selecting initial cluster centroids, assigning each observation to the closest centroid to form segments, recomputing centroids to improve the segmentation, and repeating the process until convergence. The algorithm always converges but may take longer for large datasets. The choice of distance measure, such as squared Euclidean distance or Manhattan distance, significantly impacts the segmentation solution. It is essential to repeat the algorithm with different initializations to find the best segmentation solution and assess stability before finalizing the market segments.

  - **“Improved” k-Means**

    The k-means clustering algorithm has been improved by using smart starting values instead of randomly drawing k consumers from the data set. This approach avoids the problem of the algorithm getting stuck in a local optimum, which is a good solution but not the best possible solution.

  - **Hard Competitive Learning**

    Hard competitive learning, also known as learning vector quantization, presents a variation from the standard k-means algorithm in the way segments are extracted. While both methods aim to minimize the sum of distances from each data point to its closest representative (centroid), they differ in their approach. In k-means, all data points are considered in each iteration to update the segment representatives, whereas hard competitive learning randomly selects a single data point and adjusts the closest segment representative towards this point. This procedural distinction can lead to different segmentation outcomes, even with the same initial conditions. It is possible for hard competitive learning to find the global optimal solution while k-means may converge to a local optimum, or vice versa.

- Neural Gas and Topology Representing Networks

  In summary, Neural Gas and Topology Representing Networks provide alternative clustering methods that offer more flexibility and insight into the segmentation process compared to traditional algorithms like k-means. These approaches can lead to different segmentation solutions and are valuable tools for exploratory data analysis in market segmentation.

- Self-Organizing Maps

  Self-organizing maps (SOMs), also known as Kohonen maps, are a variation of hard competitive learning used for clustering and visualization. In SOMs, segment representatives (centroids) are positioned on a regular grid, typically rectangular or hexagonal. The algorithm works by selecting a random consumer from the dataset, moving the closest representative towards them, and adjusting neighboring representatives as well. This process is repeated multiple times until a final solution is reached, with adjustments becoming smaller over iterations. One key advantage of SOMs is the structured numbering of market segments aligned with the grid, providing a systematic way to interpret the results. However, this structured positioning can lead to larger distances between segment members and representatives compared to other clustering algorithms. Studies comparing SOMs and topology representing networks with standard algorithms like k-means have been conducted for market segmentation applications.

- Neural Network

  Auto-encoding neural networks, particularly the single hidden layer perceptron, are a unique clustering method that uses three layers: input, hidden, and output. The hidden layer processes input data through weighted linear combinations to minimize the Euclidean distance between inputs and outputs. The outputs are weighted combinations of hidden nodes, with coefficients determining the relationships between hidden and output nodes. The network is trained to predict inputs accurately and accurately represent segment centroids. Unlike traditional clustering methods like k-means, neural network clustering offers fuzzy segmentation with membership values ranging between 0 and 1, allowing for more nuanced segment assignments.

- **Hybrid Approaches**

  Hybrid segmentation approaches aim to combine hierarchical and partitioning algorithms to leverage the strengths of each method while compensating for their weaknesses. Hierarchical cluster algorithms offer the advantage of not requiring the upfront specification of the number of segments and providing visual representations through dendrograms. However, they can be memory-intensive and challenging to interpret with large sample sizes. On the other hand, partitioning clustering algorithms have minimal memory requirements and are suitable for large datasets but necessitate the pre-specification of the segment number and do not facilitate tracking changes in segment membership across different solutions. The hybrid approach involves initially running a partitioning algorithm to handle datasets of any size and extracting a larger number of segments. Subsequently, only the segment centers and sizes are retained for input into hierarchical cluster analysis, enabling the determination of the optimal number of segments in a more manageable dataset.

- **Two Step Clustering**

  In the context of market segmentation analysis, the Two-Step Clustering method is a hybrid approach that combines partitioning and hierarchical clustering algorithms to segment data effectively. This method involves two main steps:

  - Initial Partitioning: In the first step, a partitioning algorithm is applied to the dataset. This algorithm does not require the upfront specification of the number of segments and can handle datasets of any size efficiently. However, instead of extracting the desired number of segments, a larger number of segments is initially generated.
  - Hierarchical Cluster Analysis: In the second step, the data is transformed by retaining only the centers (centroids) of the resulting segments and their sizes. This reduced dataset, containing representative information of each segment, is then used as input for hierarchical cluster analysis. Hierarchical clustering algorithms do not require the upfront specification of the number of segments and allow for the visualization of similarities between segments through dendrograms.

  By combining these two steps, the Two-Step Clustering method overcomes the limitations of each individual approach. The initial partitioning step efficiently handles large datasets, while the subsequent hierarchical cluster analysis benefits from a reduced dataset size and representative segment information. This hybrid approach enables data analysts to determine the optimal number of segments in a more manageable dataset, leading to more accurate and insightful market segmentation results.

- **Bagged Clustering**

  Bagged clustering is a technique used in market segmentation analysis that involves applying clustering algorithms to multiple bootstrap samples of the original dataset and then combining the results to improve the overall segmentation quality.

- **Model Based Methods**

Distance-based methods have traditionally been used in market segmentation analysis, but model-based methods have emerged as an alternative approach. Model-based methods, such as finite mixture models, have gained significant interest among marketing researchers and consultants. These methods offer a different way of extracting market segments by assuming that each segment has a certain size and specific characteristics unique to its members. The exact segment sizes and characteristics are unknown and are determined based on the empirical data.

Finite mixture models involve estimating parameters such as segment sizes and segment-specific characteristics using methods like maximum likelihood estimation or Bayesian inference. Once these parameters are estimated, consumers in the dataset can be assigned to segments based on the probability of belonging to each segment. Selecting the appropriate number of segments is a crucial aspect of model-based methods, and information criteria like AIC, BIC, and ICL are commonly used to guide this decision.

While finite mixture models may seem complex, they offer the advantage of capturing intricate segment characteristics and can be extended in various ways to accommodate different data characteristics. The literature on finite mixture models provides terminology such as mixture components, prior probabilities, and posterior probabilities to describe market segments, segment sizes, and consumer segment assignments, respectively.

Finite Mixtures of Distribution

- **Normal Distributions**

  Finite mixture models are versatile tools for market segmentation, especially when dealing with metric data. One popular approach is the mixture of multivariate normal distributions, which can effectively capture the covariance structure between variables. This model is suitable for scenarios where variables are correlated, such as physical measurements on humans or pricing in competitive markets. The parameters of the model include mean vectors and covariance matrices for each segment, with the number of parameters growing quadratically with the number of segmentation variables. In practice, fitting a mixture of normal distributions can be done using tools like the R package mclust(), which offers various covariance models to choose from. These models range from spherical (equal or varying volume) to ellipsoidal (equal or varying volume, shape, and orientation). The choice of covariance structure impacts the number of parameters to estimate, with more complex structures requiring larger sample sizes for reliable results.

  Model selection is crucial in mixture modeling, and criteria like the Bayesian Information Criterion (BIC) can help determine the optimal number of segments and covariance model. The BIC values can guide the selection of a suitable model, balancing model complexity and goodness of fit. Visualizations like uncertainty plots and BIC plots can provide insights into the segmentation results and the quality of the chosen model. Overall, finite mixture models offer a flexible and powerful approach to market segmentation, particularly for analyzing metric data with complex relationships between variables.

- Binary Distribution

  The section discusses the process of extracting segments using mixture models in market segmentation analysis. It starts by illustrating the classification plot of a mixture of normal distributions for the Australian travel motives dataset, showing the negative correlation between two variables due to different respondent groups interested in specific activities. The section then delves into fitting a mixture of binary distributions to the winter activities dataset of Austrian tourists, highlighting the association between alpine skiing and sight-seeing activities. Furthermore, the section explains the use of the R package flexmix for fitting mixture models, such as mixtures of regression models and binary distributions. It details the steps involved in fitting a mixture model with the EM algorithm, including specifying the number of segments, running multiple random restarts, and selecting the best model based on information criteria like BIC and AIC. The output from the fitting process provides insights into the convergence of the algorithm, the number of segments, and the probabilities characterizing each segment.

  Overall, the section emphasizes the importance of using mixture models in market segmentation analysis to uncover distinct segments within the data and understand the associations between variables across different segments.

Extensions and Variations

Finite mixture models offer a versatile and flexible approach to market segmentation compared to distance-based methods. These models can accommodate various data characteristics by allowing the use of different statistical models to describe market segments. Key points regarding finite mixture models include:

- **Model Flexibility**:

  Finite mixture models can utilize any statistical model to define market segments, making them adaptable to different types of data. For instance, mixtures of normal distributions can be used for metric data, mixtures of binary distributions for binary data, and mixtures of multinomial distributions or multinomial logit models for nominal variables.

- **Handling Ordinal Variables:**

  Ordinal variables pose challenges due to potential response style effects. Mixture models can disentangle these effects from content-specific responses, aiding in the extraction of meaningful market segments.

- **Incorporating Preferences:**

  Mixture models, in conjunction with conjoint analysis, enable the consideration of differences in preferences among consumers, enhancing the understanding of market segmentation.

- **Segment Variation:**

  There is a debate in segmentation literature regarding whether to model consumer differences continuously or through distinct segments. An extension to mixture models, known as mixture of mixed-effects models or heterogeneity model, acknowledges the existence of distinct segments while allowing for variation within segments.

- **Dynamic Segmentation:**

  Mixture models can be applied to data with repeated observations over time to cluster time series and extract groups of similar consumers. Dynamic latent change models, such as Markov chains, can track changes in brand choice and buying decisions over time.

- **Descriptor Variables:**

  Mixture models can incorporate both segmentation and descriptor variables. Descriptor variables help model differences in segment sizes, reflecting variations in segment composition based on specific characteristics like age or income.

Overall, finite mixture models offer a powerful tool for market segmentation, allowing for nuanced analysis of consumer behavior and preferences across various data types and scenarios.

Algorithms with Integrated Variable Selection

Algorithms with Integrated Variable Selection address the limitation of traditional segmentation algorithms by incorporating variable selection techniques within the segmentation process. While conventional algorithms assume all segmentation variables contribute equally to the segmentation solution, integrated variable selection algorithms recognize that some variables may be redundant or noisy. Pre-processing methods, such as the filtering approach by Steinley and Brusco, help identify and include only relevant variables above a certain threshold, enhancing the quality of segmentation results. For binary data, where individual variables may not be informative for clustering, specialized algorithms like biclustering and the Variable Selection Procedure for Clustering Binary Data (VSBD) by Brusco are employed to extract segments while simultaneously selecting suitable segmentation variables. Additionally, the factor-cluster analysis approach compresses segmentation variables into factors before segment extraction, offering a comprehensive two-step method for market segmentation analysis.

Variable Reduction: Factor-Cluster Analysis

Factor-cluster analysis is a two-step approach in market segmentation where segmentation variables are first factor analyzed, and then factor scores are used to extract market segments. This method is considered legitimate in cases where data originates from validated psychological tests with variables loading onto factors. However, using factor-cluster analysis to deal with a high number of segmentation variables relative to sample size lacks conceptual justification and results in a loss of information. Studies have shown that factor-cluster analysis may not outperform cluster analysis using raw data in identifying the correct market segment structure. It is recommended to perform cluster analysis on raw item scores for more accurate segmentation results and easier interpretation of segment profiles.

Data Structure Analysis

Market segmentation is an exploratory process, making traditional validation challenging as organizations cannot simultaneously test multiple segmentation strategies for optimal performance. Instead, validation in market segmentation focuses on assessing the reliability and stability of solutions through repeated calculations and modifications to the data or algorithms. This stability-based data structure analysis helps determine if natural, distinct market segments exist in the data. Data structure analysis offers valuable insights into the data properties, guiding methodological decisions and aiding in selecting the appropriate number of segments to extract. Various approaches like cluster indices, gorge plots, global stability analysis, and segment level stability analysis are used to analyze data structure and identify meaningful market segments.

- **Cluster Indices**

  In market segmentation analysis, data analysts rely on cluster indices to guide critical decisions, such as determining the number of market segments to extract. Cluster indices offer insights into different aspects of segmentation solutions and are categorized into internal and external cluster indices. Internal cluster indices are computed from a single segmentation solution, using information within that solution to assess segment similarity. On the other hand, external cluster indices compare similarity between two segmentation solutions, requiring an additional segmentation as input. Common measures like the Jaccard index, Rand index, and adjusted Rand index are used to evaluate the similarity of market segmentation solutions. Consistent outcomes across repeated calculations indicate stable extraction of market segments, enhancing the reliability of the segmentation process.

  - **Internal Cluster Indices**

    Internal cluster indices are used to evaluate a single segmentation solution and address two main questions: the compactness of market segments and the separation between different segments. These indices require a distance measure between observations and often involve calculating the sum of distances between segment members and their representative. For example, the sum of within-cluster distances (Wk) for a segmentation solution with k segments is computed by summing the distances between each segment member and their segment representative. In the context of the k-means algorithm, Wk decreases as the number of segments increases, typically visualized using a scree plot to identify an "elbow" point indicating the optimal number of segments.

Another internal cluster index, the Ball-Hall index (Wk / k), adjusts for the decrease in the internal cluster index with an increasing number of segments by dividing Wk by the number of segments. Internal cluster indices focus on assessing the similarity within segments (compactness) and the dissimilarity between segments (separation). The Ratkowsky and Lance index and the Calinski-Harabasz index are recommended internal cluster indices for variable selection and segmentation, respectively.

While internal cluster indices provide valuable insights into market segmentation solutions, they may not always guide data analysts effectively in determining the optimal number of segments, especially in consumer data where natural segments may not exist. In such cases, external cluster indices and stability analysis can be more useful in evaluating segmentation solutions.

- **External Cluster Indices**

External cluster indices play a crucial role in evaluating market segmentation solutions by incorporating additional external information that cannot be derived solely from one segmentation solution. When comparing two segmentation solutions, the issue of label switching arises due to arbitrary segment labels, leading to challenges in assessing similarities. To address this, external cluster indices like the Jaccard index and the Rand index focus on the repeated assignment of consumers to segments rather than specific labels, providing a measure of similarity between solutions. The Jaccard index, proposed by Jaccard in 1912, quantifies the similarity between two segmentation solutions based on the occurrences of consumers being assigned to the same segments. Similarly, the Rand index, introduced by Rand in 1971, considers all possible assignments of consumers to segments to determine similarity. Both indices range from 0 to 1, where 0 indicates complete dissimilarity and 1 signifies identical solutions. However, interpreting absolute values of these indices can be challenging due to the influence of segment sizes on the results.

To address the issue of chance agreement based on segment sizes, Hubert and Arabie (1985) proposed a correction for agreement by chance, leading to the development of the adjusted Rand index. This correction normalizes the index values to account for random segment assignments, providing a more accurate measure of agreement between segmentation solutions. The adjusted Rand index is particularly important in resampling-based data structure analysis approaches, offering valuable insights into the stability and reliability of market segmentation solutions.

- **Gorge Plots**

The method to assess how well segments are separated involves calculating the similarity of each consumer to the representative of a segment using the formula $S_{ih}=e^{-d_{ih}\gamma}/\sum_{l=1}^{k}e^{-d_{il}\gamma}$, where d_{ih} is the distance between consumer i and segment representative h, and γ is a hyperparameter controlling the translation of distance differences into similarity differences. These similarity values range between 0 and 1 and sum to 1 for each consumer over all segment representatives. For partitioning methods, distances are readily available, while for model-based methods, probabilities of consumer i being in segment h are used based on the fitted mixture model. Visualizing similarity values can be done using gorge plots, silhouette plots, or shadow plots. Gorge plots display histograms of similarity values for each segment, with high values indicating proximity to the segment representative and low values indicating distance. The shape of the gorge plot ideally resembles a gorge with peaks on both ends, reflecting well-separated segments.

In market segmentation analysis, gorge plots are essential for assessing segment separation. They provide insights into the distribution of similarity values, indicating how consumers relate to segment representatives. By examining gorge plots for different segment solutions, analysts can evaluate the quality of segmentation outcomes and identify natural, reproducible, or constructive segmentations. A distinct gorge plot with peaks at low and high values suggests clear segment separation, while a less defined gorge plot indicates overlapping segments. To ensure robust analysis, gorge plots should be generated and analyzed for various segment numbers. However, to address sample randomness and streamline the process, stability analysis at the global or segment level is recommended.

- **Global Stability Analysis**

The alternative approach to data structure analysis in market segmentation involves using resampling methods to assess the global stability of segmentation solutions. By generating new data sets through resampling and extracting multiple segmentation solutions, analysts can compare the stability of these solutions across repeated calculations to identify the most replicable solution. Resampling methods, such as bootstrapping, help in understanding the structure of consumer data, which may range from distinct, well-separated natural segments to entirely unstructured data. The results from global stability analysis assist in determining the most suitable number of segments to extract from the data, with a focus on reproducible and constructive segmentation solutions. This approach provides critical insight into the data structure and guides the segmentation process effectively. Global stability analysis is crucial in determining the nature of the data structure and the likelihood of natural, reproducible, or constructive segmentation.

- **Segment Level Stability Analysis**

When selecting a segmentation solution, it is important to note that choosing the globally best solution does not guarantee that it contains the single best market segment. Relying solely on global stability analysis may result in opting for a segmentation solution with overall stability but lacking a highly stable individual segment. To mitigate this risk, it is advisable to evaluate not only the global stability of various segmentation solutions but also the stability of individual market segments within those solutions. This approach safeguards against prematurely discarding solutions that may include valuable individual segments. Since most organizations typically target a single segment, assessing segment-level stability is crucial to ensure the selection of the most suitable target segment.

  - **Segment Level Stability Within Solutions (SLS_W)**

    Dolnicar and Leisch (2017) emphasize the importance of assessing segmentation solutions at the segment level to prevent discarding potentially valuable segments within an overall market segmentation solution. This approach, known as segment level stability within solutions (SLS_W), allows for the identification of highly stable segments, such as attractive niche markets, even in solutions where other segments may be unstable. The SLS_W criterion measures how consistently a market segment with similar characteristics is identified across multiple calculations of segmentation solutions with the same number of segments. This is achieved by drawing bootstrap samples, independently calculating segmentation solutions for each sample, and evaluating the agreement across all calculations using the Jaccard index.

  - **Segment Level Stability Across Solutions (SLS_A)**

    Segment level stability across solutions (SLS_A) is a criterion proposed by Dolnicar and Leisch (2017) to assess the re-occurrence of market segments across different market segmentation solutions with varying numbers of segments. High values of SLS_A indicate natural segments in the data, rather than artificially created ones. This criterion helps identify stable market segments that exist organically, making them more attractive to organizations as they do not require subjective managerial judgement in segment construction. The SLS_A plot visually represents the stability of segments across solutions with different numbers of segments. Thick lines between segments indicate stable segments that persist across solutions, while segments with changing memberships are more likely to be artificially created. Entropy, a measure of uncertainty in a distribution, can be used as a numeric indicator of segment stability across solutions. High entropy values indicate low stability, while low entropy values suggest high stability.

Step 6: Profiling Segments

- **Identifying Key Characteristics**

  Identifying the defining characteristics of market segments with respect to the segmentation variables is the aim of profiling. Profiling consists of characterising the market segments individually, but also in comparison to the other market segments. At the profiling stage, we inspect several alternative markets segmentation solutions. This is particularly important if no natural segments exist in the data, and either a reproducible or a constructive market segmentation approach must be taken. Good profiling is the basis for correct interpretation of the resulting segments. Correct interpretation, in turn, is critical to making good strategic marketing decisions.

- **Segments Profiling with Visualization**

  Neither the highly simplified, nor the very complex tabular representation typically used to present market segmentation solutions make much use of graphics, although data visualisation using graphics is an integral part of statistical data analysis. Graphics are particularly important in exploratory statistical analysis (like cluster analysis) because they provide insights into the complex relationships between variables.

  - **Identifying Defining Characteristics of Market Segments**

    A good way to understand the defining characteristics of each segment is to produce a segment profile plot. Another option is to order segmentation variables by similarity of answer patterns. A segmentation solution presented using a segment profile plot is much easier and faster to understand than when it is presented as a table, no matter how well the table is structured. Good visualisations offer an excellent return on investment.

  - **Assessing Segment Separation**

    Segment separation can be visualised using segment separation plot. The segment separation plot depicts, for all relevant dimensions of the data space, the overlap of segments. In complex situations, segment separation plot offers data analysts and users a quick overview of the data situation and the segmentation solution.

Step 7: Describing Segments

- **Developing a complete picture of Market Segments**

  Segment profiling is about understanding differences in segmentation variables across market segments. Segmentation Variables are chosen early in the market segment analysis process. In this step market segments are described using additional information available about segment members. Good description of market segments is critical to gaining detailed insights into the nature of segments. We can study difference between market segments with respect to descriptor variables in two ways: we can use descriptive statistics including visualisations or we can analyse data using inferential statistics.

- **Using Visualisation to describe market segments**

  A wide range of charts exist for the visualisation of differences in descriptor variables. There are two basic approach for nominal or ordinal descriptor variables such as gender, level of education, country of origin and metric descriptor variables such as age.

  - **Nominal and Ordinal Descriptor Variables**

    When describing differences between market segments in one single nominal or ordinal descriptor variable, the basis for all visualisations and statistical tests is a cross-tabulation of segment membership with the descriptor variable. The easiest approach to generating a cross-tabulation is to add segment membership as a categorical variable to the data frame of descriptor variables. Mosaic plot can visualise table containing more than two descriptor variables and integrate elements of inferential statistics. This helps with interpretation.

  - **Metric descriptor variables**

    R package lattice provide conditional versions of most standard R plots. Conditional plots are well-suited for visualising differences between market segments using metric descriptor variables. We can use a modified version of stability across solutions (SLS) plot to trace the value of a metric descriptor variable over a series of market segmentation solutions. The modification is that additional information contained in a metric descriptor variable is plotted using different colours.

- **Testing for Segment Differences in Descriptor Variables**

  Simple Statistical tests can be used to formally test for differences in descriptor variables across segments. The simplest way to test for differences is to run a series of independent tests for each variable of interest. Chi-square test is one of the tests that we use to test the differences. In this test p-values are calculated and based on p-value we reject or don't reject the null hypothesis. P-value<0.05 is the ground to reject the null hypothesis. The other method and the most popular method for testing for significant differences in the means of more than 2 groups is Analysis of Variance (ANOVA). In this test also we calculate p-value and if it is less than 0.05, null hypothesis is rejected. The simplest way to correct p-values for multiple testing is Bonferroni correction. Bonferroni correction multiplies p-values by the number of tests computed and, as such, represents a very conservative approach.

- **Predicting Segments from Descriptor Variables**

  We use a regression model with the segment membership as categorical dependent variable, and descriptor variables as independent variables. We can use methods developed in statistics for classification, and methods developed in machine learning for supervised learning. The prediction performance indicates how well members of a market segment can be identified given the descriptor variables. We can use the normal, Poisson, binomial, and multinomial distribution for the dependent variable in generalised linear models. The binomial or multinomial distributions are necessary for classification.

- **Binary Logistic regression**

  The binomial distribution is a generalisation of the Bernoulli distribution if the variable y does not only take values 0 and 1, but represents the number of successes out of several independent Bernoulli distributed trials with the same success probability μ. The intercept in the linear regression model gives the mean value of the dependent variable. The other coefficients in the linear regression model indicate how much the mean value of the dependent variable changes if this independent variable changes while other remains unchanged.

- **Multinomial Logistic Regression**

  Multinomial Logistic Regression can fit a model that predicts each segment simultaneously. Segment extraction typically results in more than two market segments, the dependent variable is y is not binary. Rather, it is categorical and assumed to follow multinomial distribution with the logistic function as link function. The regression coefficients are arranged in matrix form. Each row contains the regression coefficients for one category of the dependent variable. Each column contains the regression coefficients for one effect of an independent variable.

- **Tree Based Model**

  Classification and Regression trees are an alternative modelling approach for predicting a binary or categorical dependent variable given a set of independent variables. Classification and regression trees are a supervised learning technique from machine learning. The advantages of classification and regression trees are their ability to perform variable selection, ease of interpretation supported by visualisations, and the straight-forward incorporation of interaction effects. Classification and regression trees work well with many independent variables. The disadvantage is that results are frequently unstable. Small changes in the data can lead to completely different trees.

Step 8: Selecting the Target Segments

- **The Targeting Decision:**

In step 5, we can see the number of segments available for detailed inspection. These segments are profiled in step 6 and described in step 7. In step 8, segments are selected for targeting. In step 2, the knock-out and attractiveness criteria have been selected and they are now applied. Before selecting the target segment, it is important to ensure that the segments available for targeting have passed the Knockout criteria test. Once the segments have passed this test, they are required to answer two questions correctly to be shortlisted for targeting decision. These questions are as follows:

1. Which market segment(s) would the organization like to target and commit to?
2. Among the organizations offering the same product, which one(s) would each segment prefer to buy from, and how likely is it that each segment would choose our organization?

After the segments have correctly answered these questions, they are shortlisted for targeting decision.

- **Market Segment Evaluation:**

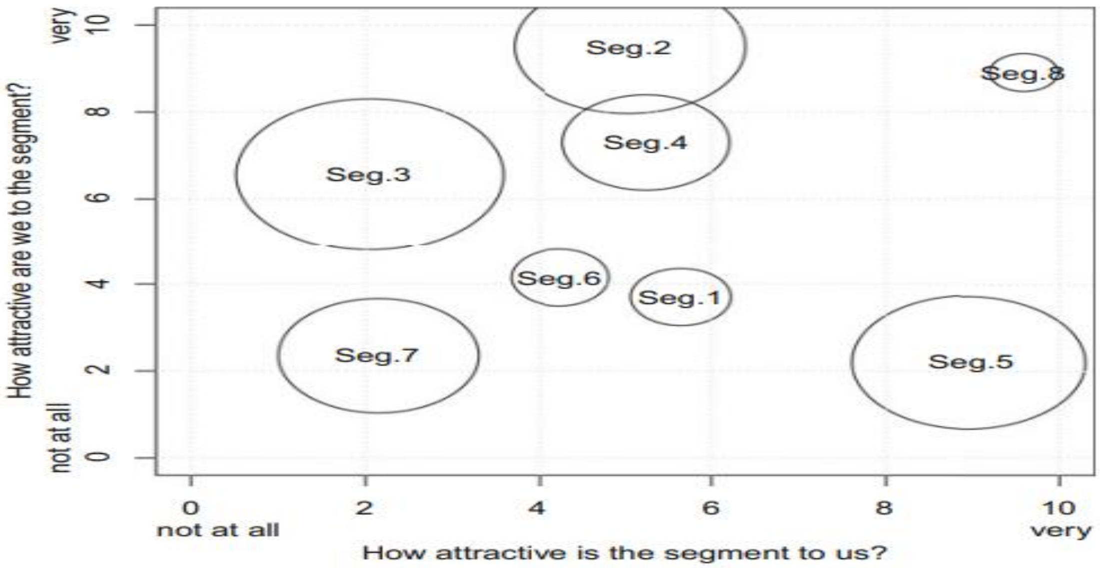
It is recommended to use a decision matrix to evaluate market segments and select one or more for targeting. The decision matrix framework to be used depends on the segmentation team, as different types of matrices have different functions that aid visualization.

The decision matrix variation has two criteria: segment attractiveness and organizational competitiveness. Since there are no best measures for these criteria, users should refer to the ideal target segment specified in step 2. Step 8 is critical, as selecting the correct target segment requires information on the attractiveness criteria for each segment, which can be obtained through grouping, profiling, and describing each market segment. To determine the location for each market segment evaluation, we multiply the weight of the segment attractiveness criterion from step 2 with the segment attractiveness criterion for each segment from steps 6 and 7. Adding the weighted values together provides the final result. Above result is plotted it looks like this it's only for reference

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 organisa-tional 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

It can also be represented in a bubble plot like below



Segments 5 and 8 are highly attractive to the organization, find the offer attractive, but have low profit potential. Segment 2 has decent profit and a favorable view. Segments 1, 4, and 6 love the organization but don't find the offer attractive. On basis of these plot user can target the Segment.

Step 9: Customising The Marketing Mix

- **Implications for Marketing Mix Decisions:**

The Marketing Mix previously relied on four important elements, known as the 4P's: Product, Price, Promotion, and Place. However, Market Segmentation is not a standalone marketing strategy, as it also involves two crucial aspects: Positioning of the organization and competition. Therefore, Segmentation is also referred to as Segmentation Targeting Positioning (STP). This means that first, the organization performs Segmentation of the market, then targets specific segments, and finally,positions its product in a way that makes it unique from competing products and meets the needs of the targeted segment.

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

The chart above illustrates how choosing a target market segment with specific positioning and competition affects different aspects of the market mix. It's important to review these aspects carefully once the target segment has been identified, to benefit the most from the market mix. Customizing the market mix to meet the needs of the target segment is essential for success. This may involve designing new products or modifying existing ones, adjusting prices, changing distribution channels, or creating new advertising campaigns. Now the market Segmentation is not done by only Considering any one 4P's but rather using all to maximize the benefits considered while selecting the Target Segment in Step 7.

- **Product:** To improve their product offerings and stay ahead in the market, organizations should consider modifying their current products instead of developing new ones. This can involve altering the packaging, introducing new names, and providing additional facilities like warranty and after-sales service support. By making these modifications, organizations can enhance the appeal of their products, attract more customers, and maintain a competitive edge in the market.
- **Price:** The organization needs to determine the appropriate price and discount for the market mix's price dimension. When developing the market mix's price dimension, the organization must carefully consider and decide on the most suitable price and discount to offer. This decision is crucial since it directly impacts the product's demand and profitability, and thus requires a comprehensive and well-informed approach.
- **Place:** The Place dimension involves deciding how to deliver the product to the customer, whether online, offline, or both, and whether to sell directly or through intermediaries like distributors, wholesalers, and retailers. These decisions have a significant impact on the product's success and the organization's ability to reach its target market.
- **Promotion:** When creating a marketing mix, it is essential to consider the promotional decision-making process. This process involves developing an advertising message that will attract the target audience. In addition, it is important to explore more effective ways of communication, such as sponsorship, personal selling, and personal relations. These tools can help to increase the effectiveness of the marketing mix and improve the overall success of the campaign.

- Github Links:
  Abhinav- https://github.com/ab04ab9752/Segmentation-code
  Akash- https://github.com/Akashkg03/McDonald-s-Market-Segmentation-Analysis
  Aun Saba-
  https://github.com/Aunsaba/Macdonalds/blob/main/Aun_Bin_Saba_Market_segmentation.ipynb