Market Segmentation Analysis

Step 01: - Deciding (not) to Segment

1.1 Implications of Committing to Market Segmentation:

- Market segmentation requires long-term commitment and is not a short-term strategy.
- Implementing market segmentation incurs costs for research, surveys, designing multiple packages and advertisements, and developing tailored communication messages.
- Organizational adjustments may be necessary, including developing new products, modifying existing products, changing pricing and distribution channels, and reorganizing around market segments.
- The decision to pursue market segmentation should be made at the highest executive level and continuously communicated and reinforced throughout the organization.
- The expected increase in sales must justify the expenses of developing and using a market segmentation strategy.
- Market segmentation is more effective when organizations organize around market segments rather than product lines.
- Effective communication and alignment are essential for successful implementation and maintenance of a market segmentation strategy.

1.2 Implementation Barriers:

I. Senior management barriers:

- Lack of leadership, commitment, and involvement from senior leadership.
- Insufficient allocation of resources for market segmentation.

II. Organizational culture barriers:

- Lack of market or consumer orientation.
- Resistance to change and new ideas.
- Lack of creative thinking and communication.

- Short-term thinking and unwillingness to make changes.
- Office politics.

III. Training and expertise barriers:

- Lack of understanding of market segmentation among senior management and the segmentation team.
- Lack of awareness of the consequences of pursuing a segmentation strategy.

IV. Lack of a qualified marketing function:

- Importance of having a formal marketing function or qualified marketing experts in the organization.
- Lack of a qualified data manager and analyst.

V. Objective restrictions and process-related barriers:

- Lack of financial resources or inability to make required structural changes.
- Lack of clear objectives, planning, and structured processes for market segmentation.
- Lack of responsibility allocation and time pressure.

VI. Acceptance and understanding barriers:

- Management's reluctance to use techniques they do not understand.
- Importance of presenting market segmentation analysis in an easily understandable manner.

If barriers cannot be overcome, it may be necessary to consider abandoning the pursuit of market segmentation. The successful implementation of market segmentation requires a resolute sense of purpose, dedication, patience, and an understanding of the challenges involved.

Barriers to implementing a market segmentation strategy include issues related to senior management, organizational culture, training and expertise, the presence of a qualified marketing function, objective restrictions, process-related challenges, and acceptance and understanding.

1.3 Checklist:

The first step in implementing a market segmentation strategy involves a checklist that includes tasks and questions. These questions serve as knock-out criteria, meaning that if the organization does not meet certain criteria, the implementation of market segmentation is likely to fail. One example is the requirement of being market-oriented. Even if a high-quality market segmentation analysis is conducted, it cannot be successfully implemented if the organization lacks a market-oriented approach.

Step 02: - Specifying the Ideal Target Segment

2.1 Segment Evaluation Criteria:

The third layer of market segmentation analysis depends primarily on user input and involvement throughout the process. User input should not be limited to just the beginning or end of the analysis but should be integrated throughout the technical aspects. The organization's contribution in Step 2 is crucial, as it guides subsequent steps, particularly data collection and selecting target segments.

In Step 2, the organization must establish two sets of segment evaluation criteria. The first set, called knock-out criteria, consists of essential features that segments must possess to be considered for targeting. The second set, known as attractiveness criteria, is used to evaluate the relative desirability of the remaining segments that meet the knock-out criteria.

While the literature does not clearly differentiate between these two types of criteria, it presents various proposed segment evaluation criteria at different levels of detail. Table 4.1 provides a selection of these criteria.

In Sections 4.2 and 4.3, segment evaluation criteria are discussed under separate headings: knock-out criteria and attractiveness criteria. The knock-out criteria are essential and non-negotiable in target segment selection. The attractiveness criteria, on the other hand, provide a diverse set of factors that the segmentation team can choose from to assess the attractiveness of potential target segments. The team must also determine the relative importance of each attractiveness criterion to the organization. While knock-out criteria eliminate certain segments automatically, attractiveness criteria are negotiated by the team and used to evaluate the overall relative attractiveness of each market segment in Step 8.

2.2 Knock-Out Criteria:

Knock-out criteria are used to assess if market segments qualify for further evaluation using attractiveness criteria. These criteria, suggested by Kotler and other authors, include homogeneity, distinctiveness, size, organizational fit, identifiability, and reachability. These criteria must be understood by senior management, the segmentation team, and the advisory committee. While most criteria are straightforward, some, like the minimum viable target segment size, require specific specification.

2.3 Attractiveness Criteria:

Table 4.1 provides a variety of segment attractiveness criteria for the segmentation team to consider. Unlike knock-out criteria, attractiveness criteria are not binary; segments are rated on a scale of attractiveness for each criterion. The overall attractiveness across all criteria determines whether a market segment is selected as a target segment in Step 8 of the market segmentation analysis.

2.4 Implementing a Structured Process:

A structured approach is recommended for evaluating market segments. The use of a segment evaluation plot, with segment attractiveness and organizational competitiveness as axes, is a popular method. The criteria for both factors need to be negotiated and agreed upon, with no more than six factors being ideal. The involvement of representatives from various organizational units is important to gather different perspectives and ensure stakeholder involvement. Selecting attractiveness criteria early on helps in data collection and makes target segment selection easier. Approximately six criteria should be chosen, each with a weight indicating its importance. Approval from the advisory committee is beneficial in finalizing the criteria.

Step 3: Collecting Data

3.1 Segmentation Variables:

Empirical data is used in both common sense and data-driven market segmentation. In common sense segmentation, a single characteristic is used as the segmentation variable to split the sample into market segments. Descriptor variables are used to describe the segments in detail. In data-driven segmentation, multiple segmentation variables are used to identify or create market segments. These segments are useful to the organization. The quality of empirical data is essential for accurately assigning individuals to segments and describing them effectively. Good data leads to better segmentation analysis and can be obtained from surveys, observations, or experimental studies. It is important to explore different data sources to ensure the data reflects actual consumer behaviour.

3.2 Segmentation Criteria:

Before extracting segments or collecting data, organizations must choose the segmentation criterion. Segmentation criteria include geographic, socio-demographic, psychographic, and behavioural factors. The decision on which criterion to use requires prior knowledge of the market and should be based on simplicity and what works best for the product or service at the least possible cost. The recommendation is to use the simplest approach that aligns with the product or service, whether it is demographic, geographic, or psychographic segmentation.

3.2.1 Geographic Segmentation:

Geographic segmentation, based on the consumer's location of residence, is often the most appropriate and simplest approach for market segmentation. It allows for targeted communication and selection of communication channels. However, it has the disadvantage of not necessarily capturing other relevant characteristics, such as consumer preferences or benefits sought. Despite its limitations, geographic information has seen a revival in international market segmentation studies aiming to extract market segments across geographic boundaries. Such studies require meaningful segmentation variables across different regions and careful consideration of cultural biases. An example is the extraction of market segments of mobile phone users among young customers across national borders.

3.2.2 Socio-Demographic Segmentation:

Socio-demographic segmentation criteria, such as age, gender, income, and education, are commonly used in certain industries like luxury goods, cosmetics, baby products, retirement villages, and tourism resorts. They offer the advantage of easy determination of segment membership and can sometimes explain specific product preferences. However, socio-demographic criteria alone may not provide sufficient market insight for optimal segmentation decisions, as they only account for a small percentage of consumer behavior variance. Values, tastes, and preferences are considered more influential in consumers' buying decisions according to Yankelovich and Meer (2006).

3.2.3 Psychographic Segmentation:

Psychographic segmentation involves grouping people based on their beliefs, interests, preferences, aspirations, or benefits sought when making a purchase. It is a more complex approach compared to geographic or socio-demographic segmentation because it requires multiple variables to capture the psychological dimension of interest. Psychographic segmentation provides insights into the underlying reasons for consumer behaviour and is commonly used in areas like tourism. However, it is more challenging to determine segment memberships for consumers, and the effectiveness of psychographic segmentation relies on the reliability and validity of the measures used to capture the psychological dimensions.

3.2.4 Behavioural Segmentation:

Behavioural segmentation involves grouping individuals based on their actual behaviours or reported behaviours, such as prior experience with a product, purchase frequency, amount spent, and information search behaviour. Using behavioural data as segmentation variables allows for segmentation based on the most relevant similarity in behaviour. It eliminates the need for developing measures for psychological constructs and focuses on real actions. However, obtaining behavioural data may not always be readily available, particularly when including potential customers who have not yet made a purchase.

3.3 Data from Survey Studies

3.3.1 Choice of Variables:

The selection of variables is crucial for the quality of market segmentation solutions. In data-driven segmentation, relevant variables must be included while unnecessary variables should be avoided. Including unnecessary variables can lead to respondent fatigue and make the segmentation process more difficult. Noisy variables, which do not contribute relevant information, can interfere with extracting optimal market segments. Careful questionnaire development and variable selection are necessary to avoid noisy variables.

3.3.2 Response Options:

The response options provided to respondents in surveys determine the scale of the data for subsequent analysis. Binary or metric response options are preferred for segmentation analysis as they facilitate distance measures and statistical procedures. Ordinal data, generated from ordered response options, pose challenges for distance measures. Using binary or metric options is recommended to avoid complications in data-driven segmentation analysis.

3.3.3 Response Styles:

The response options provided to respondents in surveys determine the scale of the data for subsequent analysis. Binary or metric response options are preferred for segmentation analysis as they facilitate distance measures and statistical procedures. Ordinal data, generated from ordered response options, pose challenges for distance measures. Using binary or metric options is recommended to avoid complications in data-driven segmentation analysis.

3.3.4 Sample Size:

The sample size plays a crucial role in segmentation analysis. Insufficient sample size makes it difficult to determine the correct number and nature of market segments. Adequate sample size allows for more accurate segmentation results. While specific guidelines are limited, a rule of thumb suggests a sample size of at least 2p (or five times 2p), where p represents the number of segmentation variables.

The research conducted by Qiu and Joe (2015) provided a sample size recommendation for constructing artificial data sets to study the performance of clustering algorithms. According to their findings, in the case of equal cluster sizes, the sample size should be at least ten times the number of segmentation variables multiplied by the number of segments in the data (10 * p * k, where p is the number of segmentation variables and k is the number of segments). If the segments are unequally sized, the smallest segment should have a sample size of at least 10 * p.

Dolnicar et al. (2014) further explored the impact of sample size on the correctness of segment recovery using artificial data resembling real-world tourism segmentation studies. They measured the correctness of segment recovery using the adjusted Rand index, which assesses the alignment between two segmentation solutions. Their results demonstrated that increasing the sample size improves the accuracy of segment identification. However, they also found that the greatest improvement is achieved by increasing small samples, and the marginal benefit decreases as the sample size increases. Based on their findings, they recommend a sample size of at least 60 * p for typical scenarios and 70 * p for more challenging artificial data scenarios.

In a subsequent study, Dolnicar et al. (2016) expanded on this research by considering additional factors that affect sample size requirements. They examined market characteristics (such as the number and size of segments and the extent of segment overlap) and data characteristics (such as sampling error, response biases, low data quality, different response options, inclusion of irrelevant items, and correlation between blocks of items). Their findings showed that larger sample sizes generally improve the ability of algorithms to identify correct segmentation solutions. However, the degree of improvement varied depending on the specific market and data characteristics. Some challenging characteristics, such as high correlation between segmentation variables, could not be adequately compensated for by increasing the sample size.

Based on the comprehensive analysis, Dolnicar et al. (2016) recommend having a sample size of at least 100 respondents per segmentation variable to enable accurate segment extraction if segments naturally exist in the data. They also emphasize the importance of collecting high-quality, unbiased data for market segmentation analysis.

In summary, to ensure reliable market segmentation results, it is important to have a sample size that is sufficiently large and meets the requirements specific to the study, including the number of segmentation variables and the characteristics of the market and data being analyzed

3.4 Data from Internal Sources:

Organizations have access to significant amounts of internal data that can be utilized for market segmentation analysis. This includes data such as scanner data from grocery stores, booking data from airline loyalty programs, and online purchase data. The strength of internal data lies in the fact that it represents actual consumer behavior rather than self-reported behavior or intentions, which can be affected by memory issues and response biases. Additionally, internal data is usually readily available as it is automatically generated and stored by organizations. However, a potential danger of using internal data is that it may be biased towards existing customers, lacking information about potential future customers with different consumption patterns.

3.5 Data from Experimental Studies:

Experimental data can serve as another valuable source for market segmentation analysis. These data can originate from field or laboratory experiments, such as tests measuring consumer responses to specific advertisements. The response to advertisements can be used as a segmentation criterion. Experimental data can also result from choice experiments or conjoint analyses, where consumers are presented with carefully designed stimuli featuring different product attributes and attribute levels. Consumers then express their preferences among the presented options. Conjoint studies and choice experiments provide insights into how different attributes and attribute levels influence consumer choice and can be used as segmentation criteria.

Step 4: Exploring Data

After data collection, exploratory data analysis cleans and – if necessary – pre-process the data. This exploration stage also offers guidance on the most suitable algorithm for extracting meaningful market segments.
Data exploration helps to (1) identify the measurement levels of the variables; (2) investigate the univariate distributions of each of the variables; and (3) assess dependency structures between variables.

Data Cleaning

The first step before commencing data analysis is to clean the data. This includes checking if all values have been recorded correctly, and if consistent labels for the levels of categorical variables have been used. For many metric variables, the range of plausible values is known in advance. levels of categorical variables can be checked to ensure they contain only permissible values.

Descriptive Analysis

Being familiar with the data avoids misinterpretation of results from complex analyses. Descriptive numeric and graphic representations provide insights into the data. Statistical software packages offer a wide variety of tools for descriptive analysis. This command returns the range, the quartiles, and the mean for numeric variables. For categorical variables, the command returns frequency counts. The command also returns the number of missing values for each variable.

Helpful graphical methods for numeric data are histograms, boxplots and scatter plots. Bar plots of frequency counts are useful for the vizualisation of categorical variables. Mosaic plots illustrate the association of multiple categorical variables. We explain mosaic plots where we use them to compare market segments.

Histograms vizualise the distribution of numeric variables. They show how often observations within a certain value range occur. Histograms reveal if the distribution of a variable is unimodal and symmetric or skewed. To obtain a histogram, we first need to create categories of values.

We call this binning. The bins must cover the entire range of observations, and must be adjacent to one another. Usually, they are of equal length. Once we have created the bins, we plot how many of the observations fall into each bin using one bar for each bin. We plot the bin range on the *x*-axis, and the frequency of observations in each bin on the *y*-axis.

Plotting density estimates allows us to superimpose probability density functions of parametric distributions. This scaling is in general viewed as the default representation for a histogram.

We can avoid selecting bin widths by using the *box-and-whisker* plot or boxplot. The boxplot is the most common graphical vizualisation of unimodal distributions in statistics. It is widely used in the natural sciences, but does not enjoy the same popularity in business, and the social sciences more generally. The simplest version of a boxplot compresses a data set into minimum, first quartile, median, third quartile and maximum.

Pre-Processing

Categorical Variables

Two pre-processing procedures are often used for categorical variables. One is merging levels of categorical variables before further analysis, the other one is converting categorical variables to numeric ones.

Merging levels of categorical variables is useful if the original categories are too differentiated (too many).

Many methods of data analysis make assumptions about the measurement level or scale of variables. The distance-based clustering methods presented assume that data are numeric, and measured on comparable scales. Sometimes it is possible to transform categorical variables into numeric variables.

Ordinal data can be converted to numeric data if it can be assumed that distances between adjacent scale points on the ordinal scale are approximately equal.

Another ordinal scale or multi-category scale frequently used in consumer surveys is the popular agreement scale which is often – but not always correctly – referred to as Likert scale (Likert 1932). Typically items measured on such a multicategory scale are bipolar and offer respondents five or seven answer options.

The verbal labelling is usually worded as follows: STRONGLY DISAGREE, DISAGREE, NEITHER AGREE NOR DISAGREE, AGREE, STRONGLY AGREE. The assumption is frequently made that the distances between these answer options are the same. If this can be convincingly argued, such data can be treated as numerical.

Binary answer options are less prone to capturing response styles, and do not require data pre-processing. Pre-processing inevitably alters the data in some way. Binary variables can always be converted to numeric variables, and most statistical procedures work correctly after conversion if there are only two categories.

Numeric Variables

To balance the influence of segmentation variables on segmentation results, variables can be standardized. Standardizing variables means transforming them in a way that puts them on a common scale.
The default standardization method in statistics subtracts the empirical mean and divides by the empirical standard deviation *s*:

Principle Component Analysis:

Principal components analysis (PCA) transforms a multivariate data set containing metric variables to a new data set with variables – referred to as principal components – which are uncorrelated and ordered by importance. The first variable (principle component) contains most of the variability, the second principle component contains the second most variability, and so on. After transformation, observations (consumers) still have the same relative positions to one another, and the dimensionality of the new data set is the same because principal components analysis generates as many new variables as there were old ones.

Principal components analysis works off the covariance or correlation matrix of several numeric variables. If all variables are measured on the same scale, and have similar data ranges, it is not important which one to use. If the data ranges are different, the correlation matrix should be used (which is equivalent to standardizing the data).

In most cases, the transformation obtained from principal components analysis is used to project high-dimensional data into lower dimensions for plotting purposes. In this case, only a subset of principal components are used, typically the first few because they capture the most variation. The first two principal components can easily be inspected in a scatter plot. More than two principal components can be visualised in a scatter plot matrix.

We interpret the output as follows: for each principal component (PC), the matrix lists standard deviation, proportion of explained variance of the original variables, and cumulative proportion of explained variance. The latter two are the most important pieces of information.

The fact that the first few principal components do not explain much of the variance indicates that all the original items (survey questions) are needed as segmentation variables. They are not redundant. They all contribute valuable information. From a projection perspective, this is bad news because it is not easy to project the data into lower dimensions. If a small number of principal components explains a substantial proportion of the variance, illustrating data using those components only gives a good visual representation of how close observations are to one another.

Sometimes principal components analysis is used for the purpose of reducing .The number of segmentation variables before extracting market segments from consumer data. This idea is appealing because more variables mean that the dimensionality of the problem the segment extraction technique needs to manage increases, thus making extraction more difficult and increasing sample size requirements . Reducing dimensionality by selecting only a limited number of principal components has also been recommended in the early segmentation literature (Beane and Ennis 1987; Tynan and Drayton 1987), but has been since shown to be highly problematic.

If all principal components would be used, the same data would be but because typically only a small subset of resulting components is used, a different space effectively serves as the basis for extracting market segments. While using a subset of principal components as segmentation variables is therefore not recommended, it is safe to use principal components analysis to explore data, and identify highly correlated variables. Highly correlated variables will display high loadings on the same principal components, indicating redundancy in the information captured by them. Insights gained from such an exploratory analysis can be used to remove some of the original – redundant – variables from the segmentation base. This approach also achieves a reduction in dimensionality, but still works with the original variables collected.

STEP-5 : Extracting Segments

Grouping Consumers: The first task in extracting segments is to group consumers based on similarities. This involves identifying patterns and similarities among individuals in the dataset.

Distance-Based Methods:

• Distance Measures: Distance-based methods calculate the similarity or dissimilarity between individuals using various distance metrics, such as Euclidean distance or Manhattan distance. These measures determine how close or far apart individuals are in the data space.

• Hierarchical Methods: Hierarchical methods create a hierarchical structure of clusters by iteratively merging or splitting clusters based on their similarity. They can be agglomerative (bottom-up) or divisive (top-down) in nature.

• Partitioning Methods: Partitioning methods aim to partition the data into a pre-defined number of clusters, where each individual belongs to a single cluster. Popular algorithms include k-means and k-medoids.

• Hybrid Approaches: Hybrid approaches combine multiple clustering techniques to improve the accuracy and effectiveness of segment extraction. These methods may incorporate both hierarchical and partitioning methods or combine distance-based and model-based approaches.

Model-Based Methods:

• Finite Mixtures of Distributions: Model-based methods assume that the data is generated from a mixture of probability distributions. Finite mixtures of distributions identify the optimal number of clusters and estimate the parameters of each distribution to assign individuals to specific segments.

• Finite Mixtures of Regressions: This approach extends the concept of finite mixtures to regression models. It allows for the identification of subpopulations with different regression patterns, enabling more precise segment extraction.

• Extensions and Variations: Model-based methods continue to evolve with new extensions and variations being developed to improve the accuracy and flexibility of segment extraction. These advancements include incorporating constraints, incorporating prior knowledge, and handling high-dimensional data.

Algorithms with Integrated Variable Selection:

• Bi-clustering Algorithms: Bi-clustering algorithms simultaneously cluster individuals and variables, identifying subsets of variables that are relevant to specific segments. These algorithms are useful when different subsets of variables are informative for different segments.

• Variable Selection Procedure for Clustering Binary Data (VSBD): This approach focuses on clustering binary data and incorporates a variable selection procedure to identify the most relevant variables for clustering.

• Variable Reduction: Factor-Cluster Analysis: Factor-cluster analysis combines factor analysis and cluster analysis to identify a reduced set of variables that capture the underlying dimensions of the data and are useful for segment extraction.

Data Structure Analysis:

• Cluster Indices: Cluster indices assess the quality and structure of the resulting clusters. They provide information about the separation, compactness, and density of the clusters.

• Gorge Plots: Gorge plots visually represent the hierarchical structure of the data, showing how clusters merge or split at different levels of similarity.

• Global Stability Analysis: Global stability analysis evaluates the stability of the clustering solution across multiple random initializations. It assesses the robustness and consistency of the segments.

• Segment Level Stability Analysis: Segment level stability analysis examines the stability of individual segments over time or across different samples. It helps determine the reliability and durability of the identified segments.

In Step 5, various methods are applied to extract segments from the data, including distance-based methods, model-based methods, algorithms with integrated variable selection, and data structure analysis. These techniques help identify meaningful and distinct segments within the target audience, which will guide further analysis and development of tailored marketing strategies for each segment.

<u>STEP-6 : Profiling Segments</u>

● Profiling is an essential step in market segmentation to understand the characteristics of market segments.

- Profiling is not necessary in commonsense segmentation as the segments are predefined based on obvious variables like age groups.
- In data-driven segmentation, the defining characteristics of market segments are unknown until after the data analysis.
- Profiling involves characterizing the market segments individually and comparing them to other segments.
- Managers often have difficulty interpreting data-driven segmentation results correctly.
- Traditional approaches to profiling market segments often present results in complex tables that are difficult to interpret.
- Comparing percentages for each segmentation variable between segments and the total sample is necessary to identify defining characteristics.
- Visualizations, such as segment profile plots, are useful for interpreting segment profiles and comparing different segmentation solutions.
- Segment profile plots show how each market segment differs from the overall sample for each segmentation variable.
- Marker variables, depicted in color, are particularly characteristic for a segment in the segment profile plot.
- Visualizations make it easier to assess the usefulness of a market segmentation solution and understand the defining characteristics of segments.

Step 7: Describing Segments

7.1 Developing a Complete Picture of Market Segments:

1.Segment Profiling:

Segment profiling involves understanding the differences in segmentation variables across market segments. These variables are chosen during the early stages of market segmentation analysis, both conceptually and empirically. They form the basis for extracting market segments from data. The goal of segment profiling is to investigate the variations in segmentation variables among different segments and gain a deeper understanding of their characteristics.

2.Segment Description:

Segment description, which is similar to segment profiling, focuses on providing additional information about market segments using variables that were not used for extracting the segments. This additional information can include demographic, psychographic, socio-economic variables, media exposure, and specific product and brand attitudes or evaluations. Describing market segments helps to create a comprehensive picture of each segment and facilitates the development of a customized marketing mix.

3.Importance of Market Segment Descriptions:

Detailed market segment descriptions are crucial for gaining insights into the nature of segments and developing a customized marketing strategy. By understanding the unique characteristics of each segment, marketers can tailor their messaging, communication channels, and offerings to effectively reach and engage with the target audience. Descriptions of market segments also help identify specific opportunities and challenges associated with each segment, allowing for more informed decision-making.

7.2 Using Vizualisations to Describe Market Segments:

Visualizations play a vital role in describing market segments as they simplify the interpretation of results for both analysts and users. They provide a graphical representation of differences in descriptor variables and can integrate information on the statistical significance of these differences. Visualizations not only enhance the understanding of market segment characteristics but also help avoid over-interpretation of insignificant differences. Graphical displays are often preferred by marketing managers for their intuitive and efficient presentation of research results.

7.2.1 Nominal and Ordinal Descriptor Variables:

When describing differences between market segments in nominal or ordinal descriptor variables (such as gender or level of education), cross-tabulation is a commonly used method. This involves creating a table that shows the distribution of segment membership across the descriptor variable categories. Visualizations, such as stacked bar charts or mosaic plots, can be used to represent these cross-tabulations. They provide a clear view of segment sizes and allow for easy comparison of proportions across segments.

Benefits of Visualizing Nominal and Ordinal Descriptor Variables:

Visualizing nominal and ordinal descriptor variables help simplify the interpretation of results and enables a quick understanding of segment differences. It allows marketers to identify any distinct patterns or variations in segment characteristics, such as gender distribution, educational backgrounds, or country of origin. By visually representing these differences, marketers can make informed decisions about targeting specific segments and developing tailored marketing strategies to effectively engage with each segment's unique characteristics.

1. Gender Distribution Across Market Segments:

- The mosaic plot and stacked bar chart illustrate the distribution of gender across market segments.
- There are no significant differences in gender distribution among the six market segments.
- The proportions of male and female tourists are approximately the same across segments.
- The size of each segment is represented by the width of the columns, while the height of the rectangles indicates the proportion of men and women within each segment.

2. Income Variation Among Market Segments:

- The mosaic plot shows a moderate association between segment membership and income levels.
- Members of segment 4, characterized by cultural interests and interactions with locals, tend to have higher incomes.
- Market segment 3, which focuses on luxury, fun, and entertainment rather than price, includes fewer low-income tourists.
- Segment 6, the nature-loving segment, has fewer members with very high incomes.

3.Association Between Travel Motives and Environmental Obligation:

- The mosaic plot reveals a strong association between travel motives and the moral obligation to protect the environment.
- The moral obligation score, ranging from low to high, is derived from survey responses about environmentally friendly behaviours.

- Segment 3, seeking entertainment, has a significantly higher proportion of members with low moral obligation scores and fewer members with high moral obligation scores.
- Conversely, segment 6, motivated by nature, has a positive association with high moral obligation and a negative association with low moral obligation.

7.2.2 Metric Descriptor Variables:

The R package lattice provides conditional plots for visualizing differences between market segments using metric descriptor variables. These plots allow for the comparison of segment profiles, such as age distribution or moral obligation scores, across different segments. Histograms and parallel box-and-whisker plots are commonly used for this purpose. Histograms, shown in Figures 9.5 and 9.6, display the distribution of age and moral obligation scores for each segment. However, assessing differences between segments solely based on these plots can be challenging.

To gain further insights, a parallel box-and-whisker plot is created for age by market segment. Figure 9.7 illustrates this plot, revealing minor differences in age across segments. The median age is lower for segment 5 and higher for segment 6. Statistical testing can be applied to validate these visual observations. In a modified version of the parallel box-and-whisker plot, box widths can be made proportional to segment sizes, and 95% confidence intervals for medians can be included. Figure 9.8 demonstrates this plot, indicating that segment 5 is the smallest and segment 6 has the highest moral obligation to protect the environment. Significant differences can be inferred if the notches for different segments do not overlap.

Another visualization, the segment level stability across solutions (SLSA) plot, can trace the value of a metric descriptor variable over multiple segmentation solutions. Figure 9.9 presents this plot with different node colours representing mean moral obligation scores. It shows that the nature-loving segment consistently displays high moral obligation, followed by the acquiescence bias segment.

These visualizations provide insights into segment differences and help guide further statistical analysis.

7.3 Testing for Segment Differences in Descriptor Variables:

Statistical tests can be used to formally test for differences in descriptor variables across market segments. The chi-square test can be used to test for independence between a nominal segment membership variable and another nominal or ordinal variable. The results of the test can be visualized using a mosaic plot. For metric variables, such as age or dollars spent on accommodation, analysis of variance (ANOVA) is commonly used to test for significant differences in means between multiple market segments. The F-test is performed, and if the p-value is smaller than 0.05, it indicates that at least two segments differ in their means.

Pairwise comparisons between segments can provide more detailed information about which segments differ significantly from each other. Tukey's honest significant differences test can be used to visualize and interpret these pairwise comparisons. Additionally, p-values should be adjusted for multiple testing to control the overall error rate. Bonferroni correction and methods such as Holm's procedure or the false discovery rate procedure can be used for this purpose.

7.4 Predicting Segments from Descriptor Variables:

In this approach, we use regression models to predict market segment membership based on descriptor variables. The regression model treats segment membership as the categorical dependent variable and the descriptor variables as independent variables. This allows us to simultaneously test the differences in all descriptor variables in relation to segment membership.The prediction performance of the regression model indicates how well we can identify members of a market segment based on the descriptor variables. Additionally, we can determine which descriptor variables are critical in identifying segment membership, especially if variable selection methods are used.

The basic regression model is the linear regression model, which assumes that the dependent variable can be predicted using the independent variables. It assumes a linear relationship between the dependent variable and the independent variables and assumes that the dependent variable follows a normal distribution with a mean determined by the independent variables.In R, the lm() function is commonly used to fit a linear regression model. The formula interface in R allows us to specify the dependent variable and independent variables. In the case of categorical variables, such as segment membership, they are appropriately coded as factors. By specifying the formula correctly, we can fit a regression coefficient for each category, indicating the mean difference between that category and a reference category.

Regression coefficients in linear regression models express how much the dependent variable changes when an independent variable changes, while all other independent variables remain constant. The intercept in the model represents the mean value of the dependent variable when all independent variables are zero or at their reference levels.

While linear regression models assume a normal distribution for the dependent variable, generalized linear models (GLMs) can accommodate a wider range of distributions. GLMs are particularly useful when the dependent variable is categorical and follows a distribution other than normal.

GLMs extend the linear regression model by introducing a link function that transforms the mean value of the dependent variable to an unlimited range. This allows for modeling the relationship between the mean value and the independent variables using a linear function. Different distributions and link functions can be used in GLMs, such as the Bernoulli distribution with the logit link function for binary logistic regression or the multinomial distribution for multinomial logistic regression.

In summary, using regression models, such as linear regression or generalized linear models, allows us to predict market segment membership based on descriptor variables. These models provide insights into the critical variables for segment identification and the relationship between the independent variables and segment membership.

7.4.1 Binary Logistic Regression:

The provided text discusses the use of generalized linear models (GLMs) for regression analysis of binary data. Specifically, it focuses on the case of binary logistic regression, where the dependent variable follows a Bernoulli distribution and the logit link function is employed. In binary logistic regression, the coefficients represent the change in the log odds of success (belonging to a particular segment) associated with a unit change in the independent variable. The odds of success are the ratio of the probability of success to the probability of failure. The text explains that the interpretation of coefficients is similar to that in linear regression, but instead of indicating changes in the mean value of the dependent variable, they reflect changes in the log odds of success.

The text also mentions the use of the glm() function in R for fitting GLMs. By specifying the appropriate family (e.g., binomial) and link function (e.g., logit), the function estimates the regression coefficients and provides information on the model fit, such as degrees of freedom, deviance, and AIC.

To interpret the coefficients and their effects more easily, the text suggests using the "effects" package in R. By calculating predicted values for different levels of the independent variables while keeping other variables constant, the package facilitates the visualization and understanding of the relationships between the predictors and the predicted probabilities of belonging to a certain segment.

The text further discusses the interpretation of the intercept in binary logistic regression, which represents the value of the linear predictor when all independent variables are zero. By applying the inverse logit function to the intercept, one can obtain the predicted probability of belonging to a certain segment for a specific combination of predictor values.

Additionally, the text mentions model selection techniques to address the inclusion of irrelevant variables and potential overfitting. The "step" function in R is highlighted as an example of a stepwise procedure that evaluates the improvement in model fit based on the AIC criterion and iteratively adds or drops variables to optimize the model's performance.

Finally, the text briefly touches on the assessment of predictive performance using the "predict()" function in R. By obtaining predicted probabilities for each observation and comparing the distributions of predicted probabilities for members and non-members of a segment, one can evaluate how well the model distinguishes between the two groups.

In summary, the text provides an overview of binary logistic regression in the context of GLMs, explains the interpretation of coefficients, describes the use of the glm() function and the "effects" package in R, mentions model selection techniques, and briefly discusses the assessment of predictive performance.

7.4.2 Multinomial Logistic Regression:

Multinomial logistic regression is used when the dependent variable has more than two categories. In R, the multinom() function from the nnet package is used to fit a multinomial logistic regression model. The model is specified using a formula and a data frame.The coefficients in a multinomial logistic regression model indicate the change in log odds of being in a particular category compared to a reference category. The coefficients are arranged in a matrix, with each row representing a category of the dependent variable and each column representing an independent variable. The Anova() function can be used to assess the significance of dropping a single variable from the model. It tests if dropping any of the variables significantly reduces the model fit.

Model selection can be performed using the step() function, which implements a stepwise procedure to select the best-fitting model based on the AIC. To assess the predictive performance of the fitted model, the predicted segment membership can be compared to the observed segment membership. Predicted probabilities for each segment can be obtained using the predict() function with the argument type = "prob". Visualizations such as mosaic plots and parallel boxplots can be used to examine the distribution of predicted segment probabilities. The allEffects() function can be used to plot the predicted probabilities of segment membership based on different independent variables, allowing for interpretation of the estimated effects. Confidence bands or intervals can be included to visualize the uncertainty of the estimated probabilities.

Overall, multinomial logistic regression allows for the simultaneous prediction of multiple categories and provides insights into the relationship between independent variables and the probability of belonging to each category.

7.4.3 Tree-Based Methods:

Classification and Regression Trees (CARTs) are a supervised learning technique used for predicting binary or categorical dependent variables based on a set of independent variables. CART models offer advantages such as variable selection, interpretability through visualizations, and the incorporation of interaction effects. They work well with a large number of independent variables. However, CART models can be unstable, as small changes in the data can lead to different trees. The tree-building process in CART involves recursively partitioning the data based on independent variables to create nodes and branches. The goal is to create groups within the nodes that are as homogeneous as possible in terms of the dependent variable. The resulting tree shows the nodes and terminal nodes, where predictions can be made based on the segment memberships of consumers within each terminal node.

There are different algorithms and packages available in R for constructing trees, such as the "rpart" package by Breiman et al. and the "partykit" package by Hothorn and Zeileis. These packages provide functions like "ctree()" to fit conditional inference trees and enable visualizations of the tree models. The output of a classification tree includes information about the nodes, splits, and terminal nodes. The tree can be visualized using the "plot()" function, which provides a clear representation of the splits, terminal nodes, and proportions of segment memberships within each node.

The construction of a classification tree involves choosing the splitting criterion, determining the minimum node size, selecting the test statistic for association tests, and setting criteria for splitting. CART models can be fitted for both binary and categorical dependent variables. When fitting a tree for a categorical dependent variable with more than two categories, each level of the dependent variable represents a segment membership, and the tree is constructed based on the associations between independent variables and segment memberships.

In summary, CART models are a useful approach for predicting binary or categorical dependent variables. They offer advantages such as variable selection and interpretability through visualizations. R packages like "rpart" and "partykit" provide functions to fit and visualize classification trees, allowing for the exploration and understanding of segment memberships based on independent variables.

STEP-8 : Selecting (the) Target Segment(s)

- The decision of target segment selection significantly impacts the organization's future performance.
- A decision matrix is commonly used to visually assess segment attractiveness and competitiveness.
- The matrix helps evaluate alternative market segments and select the most suitable ones for targeting.
- The attractiveness and competitiveness of each segment are evaluated based on predefined criteria and weights assigned to them.
- The resulting values are plotted on the matrix, with the bubble size representing additional criteria like profit potential.
- The segment evaluation plot serves as a basis for discussions and refinement of the target segment decision

STEP 9 : Customizing the Marketing Mix

Implications for Marketing Mix Decisions

Market segmentation is not a standalone marketing strategy but is closely interconnected with other strategic marketing areas such as positioning and competition. The segmentation-targeting-positioning (STP) approach is often used, which suggests a sequential process. It starts with market segmentation, where segments are identified, profiled, and described. Then, targeting involves assessing segments and selecting a target segment. Finally, positioning focuses on differentiating the product from competitors and aligning it with the needs of the chosen segment.

While viewing market segmentation as the initial step in the segmentation- targeting-positioning approach is valuable to ensure integration with other strategic decisions, it's important not to strictly adhere to a linear process. Marketers may need to move back and forth between segmentation and targeting before making a long-term commitment to one or a few target segments. Flexibility in the decision-making process is crucial.

Product

The product dimension of the marketing mix and the decisions organizations need to make about product development. The key decision is to specify the product based on customer needs. This doesn't always require creating an entirely new product but often involves modifying an existing one. Other product-related decisions include naming the product, packaging, offering warranties, and providing after-sales support services.

The paragraph then uses the example of market segments obtained from the Australian vacation activities dataset to illustrate how product design or modification is influenced by target segment selection. For instance, if a destination has a rich cultural heritage and targets segment 3, which has a strong interest in visiting museums, monuments, and gardens, the product measures could involve developing a new product. This could be a "MUSEUMS, MONUMENTS & MUCH, MUCH MORE" product accompanied by an activities pass, helping members of this segment locate activities they are in terested in during vacation planning. Another opportunity for targeting this segment could be emphasizing the destination's gardens as a prominent attraction.

Price

Setting the price: Organizations must determine the price at which their product will be sold to customers. This decision involves considering factors such as production costs, desired profit margins, market demand, competition, and perceived value of the product. The price should be set in a way that maximizes profitability while remaining attractive to customers.

Deciding on discounts: Another decision related to the price dimension is whether to offer discounts on the product. Organizations may consider various types of discounts, such as promotional discounts, volume discounts, seasonal discounts, or loyalty program discounts. The decision to offer discounts depends on factors such as the competitive landscape, marketing objectives, target market characteristics, and the desired positioning of the product.

Place

It is how to distribute the product to customers. It raises questions such as whether the product should be available for purchase online, offline, or both, and whether the
manufacturer should sell directly to customers or use intermediaries like wholesalers or retailers.

The example provided relates to segment 3 and a destination with a rich cultural heritage. The survey used for market segmentation analysis also collected data on
respondents' accommodation booking preferences during their last domestic holiday. This
information is valuable because it allows the destination to align the distribution channels of the "MUSEUMS, MONUMENTS & MUCH, MUCH MORE" product with the booking
preferences of segment 3. By understanding how segment 3 prefers to book their accommodations, the destination can ensure that the product is accessible through those distribution channels.

Promotion

The marketing mix highlights typical decisions that need to be made in this area. These decisions include developing an advertising message that resonates with the target market and determining the most effective way to communicate this message. Other promotional tools mentioned are public relations, personal selling, and sponsorship.

Referring back to Segment 3, the paragraph suggests the need to identify the best information sources to reach this segment and inform them about the "MUSEUMS, MONUMENTS & MUCH, MUCH MORE" product. This can be accomplished by comparing the information sources they used during their last domestic holiday and investigating their preferred TV stations.