Market Segmentation Analysis

Part 1 - Introduction

1.1 Strategic and Tactical Marketing

The purpose of marketing is to match the genuine needs and desires of consumers with the offers of suppliers particularly suited to satisfy those needs and desires.

Marketing plan has two components : Strategic marketing plan and Tactical marketing plan.

Strategic marketing plan : The strategic plan outlines the long-term direction of an organisation, but does not provide much detail on short- term marketing action required to move in this long-term direction

Tactical Marketing Plan : The tactical marketing plan translates the long-term strategic plan into detailed instructions for short-term marketing action

Two key decisions have to be made as part of the strategic marketing planning process: which consumers to focus on (segmentation and targeting), and which image of the organisation to create in the market (positioning).

Tactical marketing planning usually covers a period of up to one year. It covers four areas: the development and modification of the product, the determination of the price , the selection of the most suitable distribution channels, and the communication and promotion of the offer.

To conclude: the importance of strategic and tactical marketing for organisational success is asymmetric. Good tactical marketing can never compensate for bad strategic marketing. Strategic marketing is the foundation of organisational success.

1.2 Definition of Market Segmentation

According to smith the definition of market segmentation is viewing a heterogeneous market (one characterised by divergent demand) as a number of smaller homogeneous markets

Consumer characteristics deemed critical to market segmentation by management are referred to as segmentation criteria.

Concentrated market strategy: Concentrating entirely on satisfying the needs of one market segment

Differentiated market strategy : Producing products for all or multiple segments separately catering to the need of corresponding segments.

Undifferentiated market strategy : Deciding not to choose market

segmentation.

1.3 The Benefits of Market Segmentation

Market segmentation forces organisations to take stock of where they stand, and where they want to be in future.

Market segmentation leads to tangible benefits, including a better understanding of differences between consumers, which improves the match of organisational strengths and consumer needs.

Market segmentation has also been shown to be effective in sales management because it allows direct sales efforts to be targeted at groups of consumers rather than each consumer individually.

1.4 The Cost of Market Segmentation

Implementing market segmentation requires a substantial investment by the organ- isation. A large number of people have to dedicate a substantial amount of time to conduct a thorough market segmentation analysis.

The evaluation of the success of the segmentation strategy, and the continuous monitoring of market dynamics (that may point to the need for the segmentation strategy to be modified) imply an ongoing commitment of resources.

Part 2 - Ten Steps of Market Segmentation

Step 1: Deciding (not) to Segment (By Sumedh Deshkar)

3.1 Implications of Committing to Market Segmentation

Implementing a market segmentation strategy requires a long-term commitment from the organisation. It involves substantial changes, investments, and costs. The strategy should only be pursued if the expected increase in sales justifies the expenses involved. Implementing market segmentation may require developing new products, modifying existing ones, adjusting pricing and distribution channels, and adapting communication with the market. Organisational structure may also need to be reconfigured to focus on market segments rather than products. The decision to pursue market segmentation should be made at the highest executive level and communicated throughout the organisation to ensure a consistent and sustained effort.

3.2 Implementation Barriers

Successful implementation of market segmentation in organisations can face several barriers. The first group of barriers relates to senior management, including a lack of leadership, commitment, and involvement in the segmentation process. Insufficient allocation of resources by senior management can also hinder implementation. The second group of barriers relates to organisational culture, such as a lack of market orientation, resistance to change, and poor communication. Lack of training and qualified marketing expertise within the organisation can also be obstacles. Objective restrictions, like limited financial resources or the inability to make necessary structural changes, can impede implementation. Process-related barriers include unclear objectives, lack of planning, and a lack of structured processes and responsibilities. The reluctance to use unfamiliar management techniques can also hinder progress. It is crucial to identify and proactively address these barriers or consider abandoning the segmentation strategy if they cannot be overcome. Successful implementation requires dedication, patience, and a willingness to address inevitable challenges.

Step 2: Specifying the Ideal Target Segment (By Sumedh Deshkar)

4.1 Segment Evaluation Criteria

The process of market segmentation analysis requires active involvement and input from the organisation throughout various stages. The organisation's contribution is not limited to providing a brief at the beginning or developing a marketing mix at the end, but rather extends throughout the entire analysis process.

In Step 2 of the analysis, the organisation needs to determine two sets of segment evaluation criteria: knock-out criteria and attractiveness criteria. Knock-out criteria are essential and non-negotiable factors that segments must possess to be considered as potential target segments. These criteria serve as filters to eliminate segments that do not meet the organisation's requirements.

On the other hand, attractiveness criteria are used to assess the relative appeal and desirability of the remaining market segments that comply with the knock-out criteria. These criteria help in evaluating the potential of each segment and its fit with the organisation's objectives.

The literature provides a wide range of proposed segment evaluation criteria, including factors such as market size, growth rate, accessibility, profitability, compatibility, competitiveness, technological factors, socio-political considerations, and more. Different authors and researchers have suggested their own sets of criteria over time.

It is important for the segmentation team to select and prioritise the specific criteria that are most relevant and important to the organisation.

By considering both the knock-out criteria and attractiveness criteria, the organisation can make informed decisions about which market segments to target and prioritise in its marketing efforts. The involvement of the organisation in defining these criteria ensures that the segmentation analysis aligns with its strategic goals and objectives.

4.2 Knock-Out Criteria

Knock-out criteria are used to determine if market segments meet essential requirements for further assessment. These criteria include homogeneity, distinctiveness, size, fit with organisational strengths, identifiability, and reachability. Segments must be similar, distinct from others, large enough, compatible with organisational capabilities, identifiable, and reachable. Senior management, the segmentation team, and the advisory committee need to understand and agree upon these criteria. While some criteria are straightforward, others, like minimum segment size, require specific specification.

4.3 Attractiveness Criteria

In addition to knock-out criteria, Attractiveness criteria are not binary but rather assess the level of attractiveness of each segment. The combination of attractiveness across all criteria determines which segments are selected as target segments in Step 8 of the market segmentation analysis.

4.4 Implementing a Structured Process

A structured process for evaluating market segments is beneficial in market segmentation analysis. The use of a segment evaluation plot, depicting segment attractiveness and organisational competitiveness, is a popular approach. The criteria for segment attractiveness and organisational competitiveness need to be negotiated and agreed upon by the segmentation team. Involving representatives from various organisational units in this process is important due to their different perspectives and their stakeholder role in the implementation of segmentation strategy. Although the segment evaluation plot cannot be completed in Step 2, selecting attractiveness criteria early on helps in data collection and facilitates target segment selection. The segmentation team should have approximately six criteria with assigned weights indicating their importance. The weights are determined through negotiation and agreement among team members, and approval from the advisory committee is preferable to incorporate diverse perspectives.

Step 3: Collecting Data (By Sumedh Deshkar)

5.1 Segmentation Variables

In market segmentation, empirical data is used to identify or create market segments and describe them in detail. In commonsense segmentation, a single characteristic, such as gender, is used as the segmentation variable to split the sample into segments. Other characteristics serve as descriptor variables to describe the segments. In data-driven segmentation, multiple segmentation variables are used to identify naturally existing or artificially created market segments. Data quality is crucial for both types of segmentation to assign individuals to the correct segment and accurately describe the segments. Empirical data can come from

various sources, including surveys, observations (such as scanner data), or experimental studies, with the preference given to data that reflects actual consumer behaviour.

5.2 Segmentation Criteria

Before extracting segments and collecting data, organisations need to choose a segmentation criterion, which refers to the nature of the information used for market segmentation. The most common segmentation criteria are geographic, socio-demographic, psychographic, and behavioural. The decision of which criterion to use requires prior knowledge about the market and cannot be easily outsourced to consultants or data analysts. The choice of segmentation criterion should be based on the specific marketing context and the simplest approach that effectively works for the product or service. It is recommended to use the least complex criterion that achieves the desired segmentation outcomes at the lowest cost.

5.2.1 Geographic Segmentation

Geographic segmentation, which uses the consumer's location of residence as the segmentation criterion, is often the most appropriate and simple approach. It allows for easy assignment of consumers to geographic units and enables targeted communication and channel selection. However, the key disadvantage is that living in the same geographic area does not necessarily indicate shared characteristics relevant to marketers, such as product preferences or benefits sought. Other criteria, such as socio-demographic factors, are often more influential in determining consumer behaviour. Nonetheless, geographic segmentation has seen a revival in international market segmentation studies, although challenges exist in finding meaningful variables across different regions and managing biases from respondents of diverse cultural backgrounds.

5.2.2 Socio-Demographic Segmentation

Socio-demographic segmentation criteria, such as age, gender, income, and education, are commonly used and can be valuable in certain industries. They are particularly relevant for products or services targeting specific income levels, gender preferences, life stages (e.g., baby products), retirement needs, or family dynamics. Socio-demographic criteria allow for easy identification of segment membership and can sometimes explain specific product preferences (e.g., having children influencing vacation choices). However, socio-demographic factors alone may not provide sufficient market insights as they only account for a small portion of consumer behaviour variance. Values, tastes, and preferences are often more influential in shaping consumers' buying decisions, making them potentially more useful for effective market segmentation.

5.2.3 Psychographic Segmentation

Psychographic segmentation involves grouping individuals based on psychological criteria such as beliefs, interests, preferences, aspirations, and benefits sought when making purchasing decisions. It is a more complex approach compared to geographic or socio-demographic segmentation because it requires multiple variables to capture the desired psychographic dimension. Benefit segmentation, pioneered by Haley, and lifestyle segmentation are popular psychographic approaches.

The advantage of psychographic segmentation is that it provides insights into the underlying reasons for consumer behaviour. For example, identifying tourists motivated by cultural exploration can help target them with relevant offerings. However, determining segment membership using psychographic criteria is more challenging, and the effectiveness of this approach relies on the reliability and validity of the measures used to assess psychographic dimensions.

5.2.4 Behavioural Segmentation

Behavioural segmentation involves identifying similarities in actual behaviours or reported behaviours to extract market segments. These behaviours can include prior product experience, purchase frequency, purchase amount, and information search behaviour. Comparisons between segmentation

criteria have shown that behavioural variables, such as behaviours reported by tourists, can be superior to geographic variables.

The key advantage of behavioural segmentation is that it directly uses the behaviour of interest as the basis for segment extraction. This approach groups people based on the behaviour that matters most, providing valuable insights. Examples include using actual consumer expenses or purchase data as segmentation variables. Behavioural data also eliminates the need to develop measures for psychological constructs.
However, obtaining behavioural data may not always be readily available, particularly when aiming to include potential customers who have not yet made a purchase. It may require considering other approaches when studying non-existing customers of an organisation.

5.3 Data from Survey Studies

5.3.1 Choice of Variables

The careful selection of variables is crucial for the quality of market segmentation solutions, whether using commonsense segmentation or data-driven segmentation.

In data-driven segmentation, all variables relevant to the segmentation criterion need to be included, while unnecessary variables should be avoided. Including unnecessary variables can lead to long and tedious questionnaires, causing respondent fatigue and lower-quality responses. Additionally, unnecessary variables increase the dimensionality of the segmentation problem without adding relevant information, making it harder to extract optimal market segments. Noisy variables, or masking variables, do not contribute to identifying the correct market segments and can hinder the segmentation algorithm's accuracy.

To avoid noisy variables, survey questions should be carefully developed, and segmentation variables should be selected thoughtfully from available survey items. It is recommended to ask all necessary and unique questions while avoiding unnecessary or redundant ones. Redundant questions, common in traditional psychometric scale development, can interfere with the ability to identify accurate market segmentation solutions. Conducting exploratory or qualitative research can provide valuable insights for developing a good questionnaire and ensuring that no critical variables are omitted.

5.3.2 Response Options

The answer options in surveys determine the type of data available for segmentation analysis. Binary and metric data are preferred for segmentation, while nominal and ordinal data can be more challenging to analyse. Binary or metric response options are ideal, as they simplify distance measurement. Visual analogue scales can provide metric data. Binary response options often outperform ordinal options. Careful consideration of answer options is necessary to ensure the data collected is suitable for segmentation analysis.

5.3.3 Response Styles

Response styles in survey data can lead to biases and affect segmentation results. These styles include tendencies to use extreme or midpoint answer options, as well as agreeing with all statements (acquiescence bias). Segmentation algorithms may not differentiate between genuine beliefs and response styles. This can result in misleading market segments. Minimising the capture of response styles is crucial in segmentation analysis. Additional analyses or the removal of respondents displaying response styles may be necessary to ensure accurate segmentation.

5.3.4 Sample Size

Sample size plays a crucial role in market segmentation analysis. Insufficient sample sizes make it difficult to determine the correct number and nature of segments. Various studies provide recommendations for sample size requirements. Formann suggests a sample size of at least 2p (or better, five times 2p), where p is the number of segmentation variables. Qiu and Joe recommend a sample size of at least $10 \cdot p \cdot k$ (where p is the number of segmentation variables and k is the number of segments). Dolnicar et al. conducted simulation studies and recommend a sample size of at least $60 \cdot p$ for typical data sets and $70 \cdot p$ for more challenging scenarios. The sample size should be even larger if there are market characteristics like unequal segment sizes or overlapping segments.

Increasing the sample size improves the correctness of segment extraction, but the marginal benefit decreases as the sample size increases. Market and data characteristics such as correlations between variables and response biases also impact segment recovery. Uncorrelated variables lead to better segment recovery, while high correlation poses challenges even with a larger sample size.

Overall, it is important to have a sufficient sample size, typically at least 100 respondents per segmentation variable, to enable accurate segment extraction. Collecting high-quality, unbiased data is also crucial for reliable market segmentation analysis.

In summary, for optimal market segmentation results, the data should include all necessary items, exclude unnecessary and correlated items, have high-quality responses, be binary or metric, be free of response styles, and have a sample size appropriate for the number of segmentation variables (recommended as 100 times the number of variables).

5.4 Data from Internal Sources

Organisations now have access to large volumes of internal data that can be utilised for market segmentation analysis. Examples include scanner data from grocery stores, booking data from airline loyalty programs, and online purchase data. The strength of such data lies in the fact that it represents actual consumer behaviour rather than self-reported information, which can be influenced by memory limitations and response biases.

Internal data is advantageous because it is automatically generated and readily available if stored in an accessible format. However, a potential danger is that the data may be biassed towards existing customers, lacking information about potential future customers who may have different consumption patterns.

5.5 Data from Experimental Studies

Experimental data, whether from field or laboratory experiments, can serve as another valuable source for market segmentation analysis. These experiments can involve testing how individuals respond to different advertisements, with their responses becoming segmentation criteria. Additionally, experimental data can arise from choice experiments or conjoint analysis, where consumers are presented with various product attributes and levels and asked to indicate their preferences. These studies provide insights into the influence of different attributes and levels on consumer choices, which can be utilised as segmentation criteria.

Step 4: Exploring Data (By Shagun Sirohi)

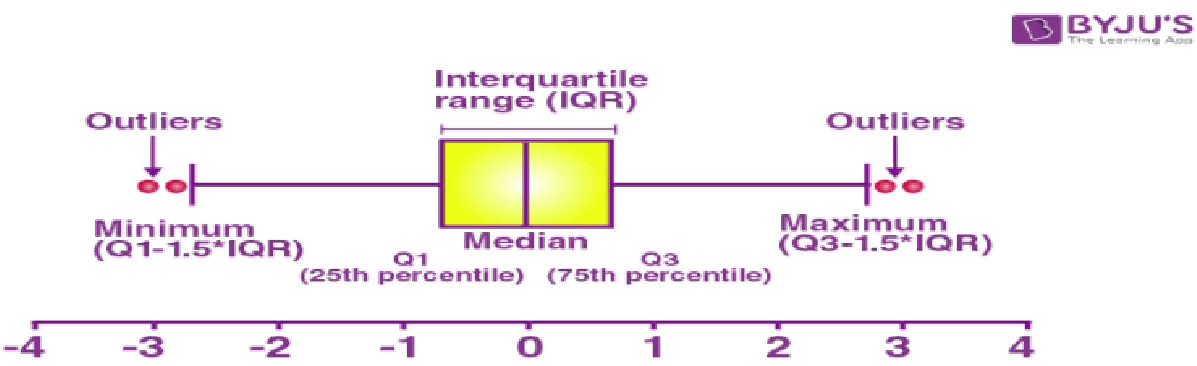
After data collection, exploratory data analysis cleans and – if necessary – preprocesses the data. This exploration stage also offers guidance on the most suitable algorithm for extracting meaningful market segments. Data CleaningThe first step before commencing data analysis is to clean the data. This 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. For example, age (in years) can be expected to lie between 0 and 110. It is easy to check whether any implausible values are contained in the data, which might point to errors during data collection or data entry. Similarly, levels of categorical variables can be checked to ensure they contain only permissible values. For example, gender typically has two values in surveys: female and male. Unless the questionnaire did offer a third option, only those two should appear in the data. Any other values are not permissible, and need to be corrected as part of the data cleaning procedure.

```
inc2 = vac['Income2'] levels = ['<30k', '>120k', '30-60k', '60-90k', '90-120k']
inc2 = pd.Categorical(inc2, categories=levels, ordered=True) inc2 =
inc2.rename_categories(['<30k', '30-60k', '60-90k', '90-120k', '>120k'])
print(inc2)
```

```
orig = vac['Income2'] new = inc2 table = pd.crosstab(orig, new)
```

Descriptive Analysis-we obtain a numeric summary of the data with command summary(). 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 visualisation of categorical variables. Mosaic plots illustrate the association of multiple categorical variables. We explain mosaic plots in Step 7 where we use them to compare market segments. Histograms visualise 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.

```
plt.hist(vac['Age'])
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.title('Histogram of Age')
plt.show()
```




```
import pandas as pd
import matplotlib.pyplot as plt
```

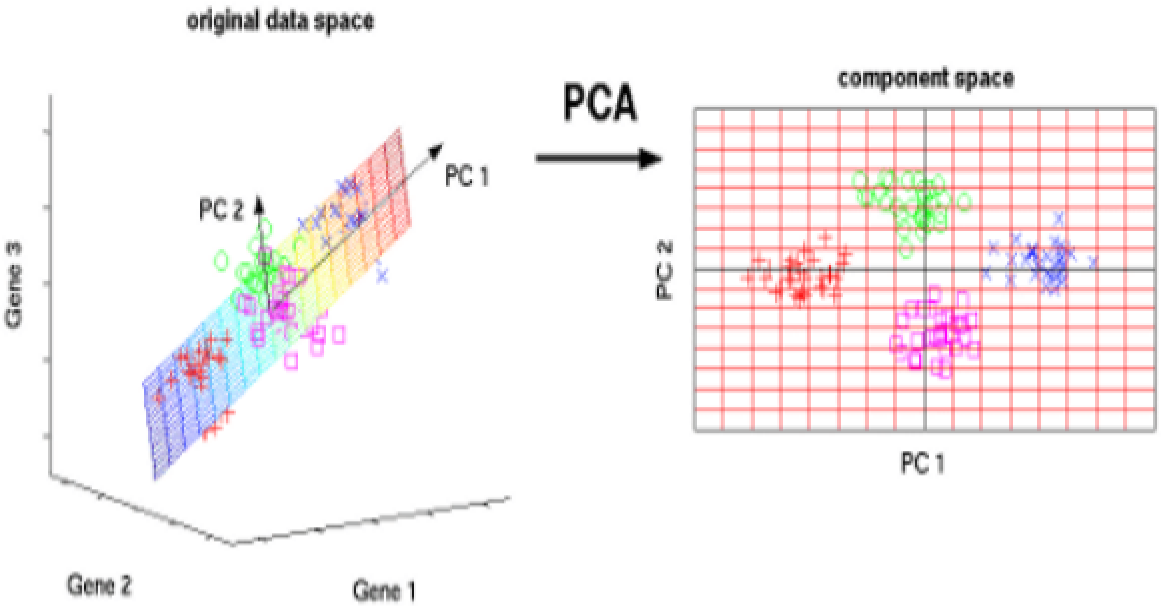
```
yes = 100 * (vac.iloc[:, 12:31] == "yes").mean()

sorted_yes = yes.sort_values()

plt.figure(figsize=(6, 8))
plt.scatter(sorted_yes, range(len(sorted_yes)), color='blue')
plt.xlabel("Percent 'yes'")
plt.xlim(0, 100)
plt.yticks(range(len(sorted_yes)), sorted_yes.index)
plt.title("Dot Chart of Percent 'yes'")
plt.show()
```

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, if it makes sense to do so. Merging levels of categorical variables is useful if the original categories are too differentiated (too many). Thinking back to the income variables, for example, the original income variable as used in the survey has the following categories: Many methods of data analysis make assumptions about the measurement level or scale of variables. The distance-based clustering methods presented in Step 5 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. This is a reasonable assumption for income, where the underlying metric construct is classified into categories covering ranges of equal length.

Numeric Variables - The range of values of a segmentation variable affects its relative influence in distance-based methods of segment extraction. If, for example, one of the segmentation variables is binary (with values 0 or 1 indicating whether or not a tourist likes to dine out during their vacation), and a second variable indicates the expenditure in dollars per person per day (and ranges from zero to $1000), a difference in spend per person per day of one dollar is weighted equally as the difference between liking to dine out or not. To balance the influence of segmentation variables on segmentation results, variables can be standardised. Standardising variables means transforming them in a way that puts them on a common scale.

Principal Components 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 basically keeps the data space unchanged, but looks at it from a different angle. 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 standardising 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.

```python
import pandas as pd
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
data_scaled = scaler.fit_transform(data)
pca = PCA()
principal_components = pca.fit_transform(data_scaled)
explained_variance_ratio = pca.explained_variance_ratio_
cumulative_variance = np.cumsum(explained_variance_ratio)
print("Explained Variance Ratio:")
print(explained_variance_ratio)
print("\nCumulative Explained Variance:")
print(cumulative_variance)
```



Step 5 - Extracting Segments (Shagun Shirohi)

Data-driven market segmentation analysis is exploratory by nature. Consumer data sets are typically not well structured. Consumers come in all shapes and forms; a two-dimensional plot of consumers' product preferences typically does not contain clear groups of consumers. Rather, consumer preferences are spread across the entire plot. The combination of exploratory methods and unstructured consumer data means that results from any method used to extract market segments from such data will strongly depend on the assumptions made on the structure of the segments implied by the method. The result of a market segmentation analysis, therefore, is determined as much by the underlying data as it is by the extraction algorithm chosen. Segmentation methods shape the segmentation solution. Many segmentation methods used to extract market segments are taken from the field of cluster analysis. In that case, market segments correspond to clusters. As pointed out by Hennig and Liao (2013), selecting a suitable clustering method requires matching the data analytic features of the resulting clustering with the context-dependent requirements that are desired by the researcher (p. 315). It is, therefore, important to explore market segmentation solutions derived from a range of different clustering methods. It is also important to understand how different algorithms impose structure on the extracted segments.

Distance-Based Methods - distance-based methods in the context of market segmentation for tourists based on vacation activity patterns. It introduces the concept of distance measures and presents three common distance measures: Euclidean distance, Manhattan distance, and asymmetric binary distance. The text explains the criteria for a valid distance measure and illustrates the differences between the measures using a fictitious data set. It further discusses the usage of the python programming language for calculating distances using the dist() function. The examples demonstrate how to calculate Euclidean distance and Manhattan distance between tourists' vacation activity profiles. The text also mentions the daisy() function from the R package "cluster," which allows for dissimilarity calculations when dealing with different variable types. Overall, the text provides an overview of distance-based methods and their application in market segmentation analysis for grouping tourists based on their vacation activity patterns.

Hierarchical Methods - Hierarchical methods are a class of distance-based methods used in market segmentation analysis to group similar observations based on their characteristics. These methods build hierarchical structures, such as dendrograms or trees, to represent the relationships between observations. This summary provides an overview of hierarchical methods in market segmentation and their implementation in Python. Hierarchical clustering is a widely used hierarchical method that recursively merges observations based on their proximity or similarity. It starts with each observation as an individual cluster and iteratively combines clusters until a single cluster containing all observations is formed. The linkage method determines how the distance between clusters is calculated, with popular options including Ward's method, complete linkage, and average linkage. In Python, the scipy.cluster.hierarchy module provides functions for performing hierarchical clustering. The linkage function computes the linkage matrix that captures the pairwise distances or dissimilarities between observations. The resulting matrix is then used to generate a dendrogram, a tree-like visualization of the hierarchical clustering structure. The dendrogram allows for the identification of clusters at different levels of similarity or dissimilarity. The code snippet presented in Python demonstrates how to perform hierarchical clustering using the Ward linkage method and visualize the resulting dendrogram. The linkage function computes the linkage matrix, while the dendrogram function creates the dendrogram plot. Additional functions from the matplotlib.pyplot module are used to set labels and titles for the plot. Overall, hierarchical methods provide a useful framework for market segmentation analysis, enabling the identification of clusters and subclusters based on the similarity of observations. The Python code snippet showcases the implementation of hierarchical clustering using the scipy library, allowing for practical application and visualization of these methods in market segmentation studies.

Partitioning Methods - Partitioning methods are a class of distance-based methods commonly used in market segmentation analysis to divide observations into distinct nonoverlapping groups. This summary provides an overview of partitioning methods in market segmentation and their implementation in Python. Partitioning algorithms, such as K-means clustering, aim to create clusters by minimizing the within-cluster variance or maximizing the between-cluster distances. These methods iteratively assign observations to clusters and update the cluster centroids until a convergence criterion is met. In Python, the scikit-learn library provides efficient implementations of partitioning methods. The KMeans class allows for performing K-means clustering on a given dataset. The number of desired clusters, K, needs to be specified. After fitting the model, the resulting clusters can be accessed using the labels_ attribute, while the centroids of each cluster are available through the cluster_centers_ attribute. The code snippet demonstrates how to perform K-means clustering in Python using the scikitlearn library. It involves importing the KMeans class, initializing an instance with the desired number of clusters, fitting the model to the data, and accessing the resulting cluster labels and centroids. Additionally, other partitioning methods such as the Fuzzy C-means (FCM) algorithm, which allows for fuzzy clustering where observations can belong to multiple clusters with different degrees of membership, can also be implemented using appropriate libraries or custom functions. Partitioning methods offer a flexible and efficient approach to market segmentation analysis, allowing for the creation of distinct clusters based on the proximity or similarity of observations. By utilizing the capabilities of Python and relevant libraries, researchers can easily implement and analyze partitioning methods to gain insights into market segmentation problems.

Model-Based Methods - Model-based methods are a class of techniques used in market segmentation analysis that aim to identify underlying patterns or structures in the data through the use of statistical models. This summary provides an overview of model-based methods in market segmentation analysis and their significance in gaining insights into consumer behavior. Model-based methods involve specifying a probabilistic model that describes the data generation process and the relationships between different variables. These models can capture complex interactions and dependencies among the observed variables and are used to infer latent variables or unobservable factors driving consumer behavior. One popular model-based method is Latent Class Analysis (LCA), which assumes the existence of distinct latent classes or segments within the population. LCA estimates the probabilities of individuals belonging to each class and assigns them to the most likely segment based on their observed characteristics. This allows for the identification of meaningful and interpretable segments with distinct preferences or behavior patterns. Another commonly used model-based method is Finite Mixture Modeling, which extends LCA to continuous variables and assumes that the data distribution is a mixture of several component distributions. This method allows for more flexibility in modeling various types of data, such as continuous, categorical, or count variables. Python provides several libraries for implementing model-based methods in market segmentation analysis. The statsmodels library offers functionalities for fitting Latent Class Models (LCMs) and estimating class membership probabilities. Additionally, the scikit-learn library provides tools for Finite Mixture Modeling using Gaussian Mixture Models (GMMs). The code snippet below demonstrates the implementation of Latent Class Analysis using the statsmodels library in Python:

```
import pandas as pd
import statsmodels.api as sm

# Assuming 'data' is a DataFrame containing the input data

# Specify the LCM model
model = sm.LatentClassModel(data)

# Fit the model
result = model.fit()

# Get the estimated class probabilities
class_probs = result.predict()

# Assign individuals to segments based on maximum probability
segments = class_probs.idxmax(axis=1)
```

Model-based methods offer a powerful approach to market segmentation analysis by providing a theoretical framework for understanding consumer behavior. These methods can uncover hidden segments, discover relationships between observed variables, and provide insights into consumer preferences and decision-making processes. The implementation of model-based methods in Python enables researchers to leverage the available tools and libraries to gain valuable insights from their data.

Algorithms with Integrated Variable Selection - Algorithms with integrated variable selection are a class of techniques used in market segmentation analysis that simultaneously perform both variable selection and clustering to identify meaningful segments based on relevant variables. This summary provides an overview of these algorithms and their significance in market segmentation. Traditional market segmentation approaches often rely on pre-determined variables, assuming that all variables are equally important for segmentation. However, in reality, not all variables may contribute equally to segment differentiation. Algorithms with integrated variable selection aim to overcome this limitation by automatically selecting the most relevant variables during the segmentation process. One popular algorithm with integrated variable selection is the Integrated Fuzzy Clustering and Variable Selection (IFCVS) algorithm. This algorithm combines fuzzy clustering, which allows for soft assignments of observations to clusters, with variable selection techniques such as genetic algorithms or information gain. It optimizes both the clustering structure and the subset of variables used for segmentation, resulting in more accurate and interpretable segments. Another commonly used algorithm is the Variable Selection Integrated Fuzzy C-Means (VSIFCM) algorithm. This method combines Fuzzy C-Means clustering with variable selection techniques to identify the most discriminative variables for segmentation. By incorporating variable selection, the algorithm enhances the quality of the resulting clusters and identifies the subset of variables that contribute the most to the segmentation process. Python provides various libraries and tools to implement algorithms with integrated variable selection in market segmentation analysis. The scikit-fuzzy library offers functionalities for fuzzy clustering, while variable selection techniques can be implemented using libraries such as scikit-learn or custom functions. The code snippet below demonstrates the implementation of the VSIFCM algorithm in Python using the scikit-fuzzy library:

```
import pandas as pd
import skfuzzy as fuzz

# Assuming 'data' is a DataFrame containing the input data

# Perform variable selection using appropriate technique
selected_variables = perform_variable_selection(data)

# Extract selected variables from the data
selected_data = data[selected_variables]

# Perform fuzzy clustering using the selected variables
centers, u, u0, d, jm, p, fpc = fuzz.cluster.cmeans(
    selected_data.T, c=3, m=2, error=0.005, maxiter=1000
)

# Assign individuals to clusters based on maximum membership
segments = u.argmax(axis=0)
```

Algorithms with integrated variable selection provide an efficient and effective approach to market segmentation by considering the importance of variables in the segmentation process. By automatically selecting relevant variables, these algorithms enhance the accuracy and interpretability of the resulting segments. The implementation of these algorithms in Python allows researchers to leverage the available tools and libraries to perform advanced market segmentation analyses.

Data Structure Analysis – Extracting market segments is inherently exploratory, irrespective of the extraction algorithm used. Validation in the traditional sense, where a clear optimality criterion is targeted, is therefore not possible. Ideally, validation would mean calculating different segmentation solutions, choosing different segments, targeting them, and then comparing which leads to the most profit, or most success in mission achievement. This is clearly not possible in reality because one organisation cannot run multiple segmentation strategies simultaneously just for the sake of determining which performs best. Cluster indices, Gorge plots, global stability analysis, and segment level stability analysis are data structure analysis techniques used in market segmentation analysis to evaluate the quality and stability of segmentation results. This summary provides an overview of these techniques and their significance in assessing the robustness and reliability of segmentation outcomes. Cluster indices are quantitative measures that assess the compactness and separation of clusters within a segmentation solution. Common cluster indices include the Dunn Index, Silhouette Score, and Calinski-Harabasz Index. These indices provide insights into the distinctiveness and quality of the identified segments, allowing for the comparison and selection of the most appropriate segmentation solution. Gorge plots, also known as scree plots or elbow plots, are graphical representations that show the relationship between the number of clusters and a clustering criterion, such as the withincluster sum of squares (WCSS) or total variance explained. Gorge plots help determine the optimal number of clusters by identifying the "elbow" or "knee" point where the incremental gain in clustering quality diminishes significantly. Global stability analysis evaluates the stability of the segmentation solution across multiple random subsamples of the data. It assesses the consistency of the clusters and examines whether the same segments consistently emerge across different subsamples. This analysis provides insights into the robustness of the segmentation solution and its generalizability to the entire dataset. Segment level stability analysis focuses on evaluating the stability of individual segments over time or across different datasets. It assesses whether the segments remain meaningful and consistent across different time periods, customer cohorts, or other relevant criteria. This analysis helps determine the stability and reliability of the identified segments, ensuring their usefulness for marketing strategies and decision-making. These data structure analysis techniques can be implemented using various statistical and visualization methods in Python. Libraries such as scikit-learn, matplotlib, and seaborn provide functions and tools for computing cluster indices, generating gorge plots, conducting global stability analysis, and performing segment level stability analysis. By employing these techniques, researchers can gain insights into the quality, robustness, and stability of segmentation solutions, enabling more informed decision-making in marketing strategies and customer targeting effort.

Step 6: Profiling Segments (By Sumedh Deshkar)

Identifying Key Characteristics of Market Segments

The aim of the profiling step in market segmentation is to understand and characterise the resulting market segments. In data-driven segmentation, where segments are derived from analysis of consumer data, profiling helps identify the defining characteristics of each segment. This is important for interpreting the segmentation results correctly and making effective marketing decisions. However, in commonsense segmentation, where predefined segments based on obvious criteria are used (e.g., age groups), profiling is not necessary. Profiling involves comparing and contrasting different segmentation solutions, especially when natural segments do not exist in the data. Data-driven segmentation solutions can be challenging to interpret, and many marketing managers struggle to understand them. Graphical statistics approaches can help make profiling less tedious and prone to misinterpretation.

Traditional Approaches to Profiling Market Segments

In the context of data-driven market segmentation using the Australian vacation motives dataset, the profiling step aims to identify and understand the defining characteristics of the extracted market segments. Profiling involves characterising each segment individually and comparing them to other segments.

Traditional approaches to profiling market segments often involve presenting large tables that provide exact percentages for each segmentation variable within each segment. However, interpreting such tables can be challenging and time-consuming, especially when dealing with multiple segments and variables. Comparing the percentages within each segment to the total and to other segments requires a substantial number of comparisons.

The given example in the book illustrates the mean values (percentages) of travel motives for each segment. By comparing the percentages, it becomes possible to identify the defining characteristics of each segment. However, without the table information, it is difficult to provide a detailed summary of the segment profiles.

It is worth noting that presenting segmentation results solely as tables can be misleading or overwhelming. Managers and clients often struggle to interpret and understand data-driven segmentation solutions, as they may view them as black boxes. Simplifying and visually representing the segment profiles using graphical statistics approaches can help improve interpretation and decision-making.

Overall, the profiling step is essential for correctly interpreting market segments and making informed strategic marketing decisions.

Segment Profiling with Visualisations

Graphics and data visualisation play a crucial role in market segmentation analysis by providing insights and intuitive interpretations. While traditional tabular representations are commonly used, they often lack the comprehensive view and ease of understanding offered by graphical techniques. Visualisations aid in exploring complex relationships between variables, monitoring developments over time, and evaluating different segmentation solutions. By presenting information in a graphical format, visualisations enhance the interpretation of segment profiles and
assist in decision-making. Their use in the data-driven market segmentation process improves the understanding of segments and supports marketers and analysts in making informed choices.

Identifying Defining Characteristics of Market Segments

Segment profile plots, visual representations of market segments, are valuable tools for understanding the characteristics of each segment. These plots depict how each segment differs from the overall sample across segmentation variables. By clustering the variables and arranging them meaningfully, the plot becomes easier to interpret. Marker variables, representing segment characteristics, are highlighted in colour, while other variables are greyed out. The segment profile plot provides a comprehensive view of each segment's attributes and allows for easy comparison with the overall sample. Comparing visualisations to traditional tabular presentations, it is evident that segment profile plots offer quicker and more intuitive interpretations, making them essential for strategic decision-making based on segmentation analysis.

Assessing Segment Separation

Segment separation plots are visualisations that depict the overlap of segments in a data space. They provide data analysts and users with a quick overview of the data situation and the segmentation solution. The plots consist of scatter plots of observations coloured by segment membership, cluster hulls indicating the shape and spread of segments, and neighbourhood graphs showing the similarity between segments.

For higher-dimensional data, projection techniques are used to create segment separation plots. Principal components analysis (PCA) is one such technique. The data is projected onto a small number of dimensions, and the resulting plot shows the separation between segments in the projected space. In Fig. 8.5, a segment separation plot using principal components 2 and 3 is shown for the Australian travel motives data set.

To enhance the readability of segment separation plots, modifications can be made, such as modifying colours, omitting observations, and highlighting only the inner area of each segment. Fig. 8.6 shows a cleaner version of the plot, where the segments are easier to interpret.

It's important to note that each segment separation plot represents a specific projection of the data, and different projections may result in different visualisations of segment separation. Therefore, conclusions about segment overlap should be based on multiple projections, not just a single one.

Overall, segment separation plots provide valuable insights into the separation and overlap of segments in a data set, helping analysts understand the characteristics and differences between segments.

Step 7 - Describing Segments (By Roshit Dahat)

7.1 : Developing a Complete Picture of Market Segments:

Segment profiling is an important aspect of market segmentation analysis. It involves understanding the differences in segmentation variables across different market segments. Segmentation variables are chosen early in the analysis process and form the basis for extracting market segments from data.

Step 7 of the analysis process, known as segment description, is similar to profiling but involves describing market segments using additional information about segment members. This includes variables like psychographic, demographic, and socio-economic factors, media exposure, and product and brand attitudes.

In the context of a data-driven market segmentation analysis using the Australian travel motives data set, profiling investigates differences between segments based on travel motives. Segment description utilizes additional variables such as age, gender, past travel behavior, media use, and expenditure patterns.

Good descriptions of market segments are crucial for gaining detailed insights and developing a customized marketing mix. By understanding the characteristics of a particular segment, targeted communication and marketing strategies can be developed. Descriptive statistics and visualizations are useful tools for studying differences between market segments based on descriptor variables.

Overall, segment profiling and description are essential for effective market segmentation and developing tailored marketing approaches.

7.2: Using Visualizations to Describe Market Segments

Visualizing differences in descriptor variables is essential for effective market segmentation analysis. There are various chart options available for this purpose. Two common approaches are discussed: one suitable for nominal and ordinal variables (e.g., gender, education level, country of origin), and another for metric variables (e.g., age, number of nights, money spent).

Using graphical statistics offers two advantages: it simplifies interpretation for both analysts and users, and it incorporates information on the statistical significance of differences, preventing the over-interpretation of insignificant variances. Graphical representations are highly valued by marketing managers for their intuitive nature and effectiveness in conveying research results. In fact, graphical displays are found to be processed more efficiently than tabular formats.

About the Australian Data Set:

Year of data collection: 2006.
Location: Australia.
Sample size: 1000.
Sample: Adult Australian residents.
Segmentation variables used in the Report:
Twenty travel motives, integer codes are 1 (for applies) and 0 (for does not apply).

- I want to rest and relax (REST AND RELAX)
- I am looking for luxury and want to be spoilt (LUXURY / BE SPOILT)
- I want to do sports (DO SPORTS)
- This holiday means excitement, a challenge and special experience to me (EXCITEMENT, A CHALLENGE)
- I try not to exceed my planned budget for this holiday (NOT EXCEED PLANNED BUDGET)
- I want to realize my creativity (REALIZE CREATIVITY)
- I am looking for a variety of fun and entertainment (FUN AND ENTERTAINMENT)
- Good company and getting to know people is important to me (GOOD COMPANY)
- I use my holiday for the health and beauty of my body (HEALTH AND BEAUTY)
- I put much emphasis on free-and-easy-going (FREE-AND-EASY-GOING)
- I spend my holiday at a destination, because there are many entertainment facilities (ENTERTAINMENT FACILITIES)
- Being on holiday I do not pay attention to prices and money (NOT CARE ABOUT PRICES)
- I am interested in the lifestyle of the local people (LIFESTYLE OF THE LOCAL PEOPLE)
- The special thing about my holiday is an intense experience of nature (INTENSE EXPERIENCE OF NATURE)
- I am looking for coziness and a familiar atmosphere (COSINESS/FAMILIAR ATMOSPHERE)
- On holiday the efforts to maintain unspoilt surroundings play a major role for me (MAINTAIN UNSPOILT SURROUNDINGS)
- It is important to me that everything is organized and I do not have to care about anything (EVERYTHING ORGANISED)
- When I choose a holiday-resort, an unspoilt nature and a natural landscape plays a major role for me (UNSPOILT NATURE/NATURAL LANDSCAPE)
- Cultural offers and sights are a crucial factor (CULTURAL OFFERS)
- I go on holiday for a change to my usual surroundings (CHANGE OF SURROUNDINGS)

The three numeric descriptor variables OBLIGATION, NEP, VACATION.BEHAVIOUR (see below) are also used as segmentation variables to illustrate the use of model based methods.

Descriptor variables used in the Report:

- Gender (FEMALE, MALE)
- Age (numeric)
- Education (numeric, minimum 1, maximum 8)
- Income (LESS THAN $30,000, $30,001 TO $60,000, $60,001 TO $90,000, $90,001 TO $120,000, $120,001 TO $150,000, $150,001 TO $180,000, $180,001 TO $210,000, $210,001 TO $240,000, MORE THAN $240,001)
- Re-coded income (<30K, 30–60 K, 60–90 K, 90–120 K, >120K)
- Occupation (CLERICAL OR SERVICE WORKER, PROFESSIONAL, UNEMPLOYED, RETIRED, MANAGER OR ADMINISTRATOR, SALES, TRADESPERSON,

SMALL BUSINESS OWNER, HOME-DUTIES, TRANSPORT WORKER, LABORER)
• State (NSW, VIC, QLD, SA, WA, TAS, NT, ACT)
• Relationship status (SINGLE, MARRIED, SEPARATED OR DIVORCED, LIVING WITH A PARTNER, WIDOWED)
• Stated moral obligation to protect the environment (OBLIGATION: numeric, minimum 1, maximum 5).
• Re-coded stated moral obligation to protect the environment (OBLIGATION2: recoded ordered factor by quartiles: Q1, Q2, Q3, Q4).
• Mean New Ecological Paradigm (NEP) scale value (NEP: numeric, minimum 1, maximum 5).
• Mean environmental friendly behavior score when on vacation (VACATION.BEHAVIOUR: numeric, minimum 1, maximum 5).

Data set vacmot (containing the three objects vacmot, vacmot6 and vacmotdesc)

7.2.1: Nominal and Ordinal Descriptor Variables

When describing differences between market segments in one single nominal or ordinal descriptor variable, the basis for all visualizations and statistical tests is a cross-tabulation of segment membership with the descriptor variable. For the Australian travel motives data set, data frame vacmotdesc contains several descriptor variables.These descriptor variables are automatically loaded with the Australian travel motives data set.
To describe market segments, we need the segment membership for all respondents. The sizes of the market segments are:

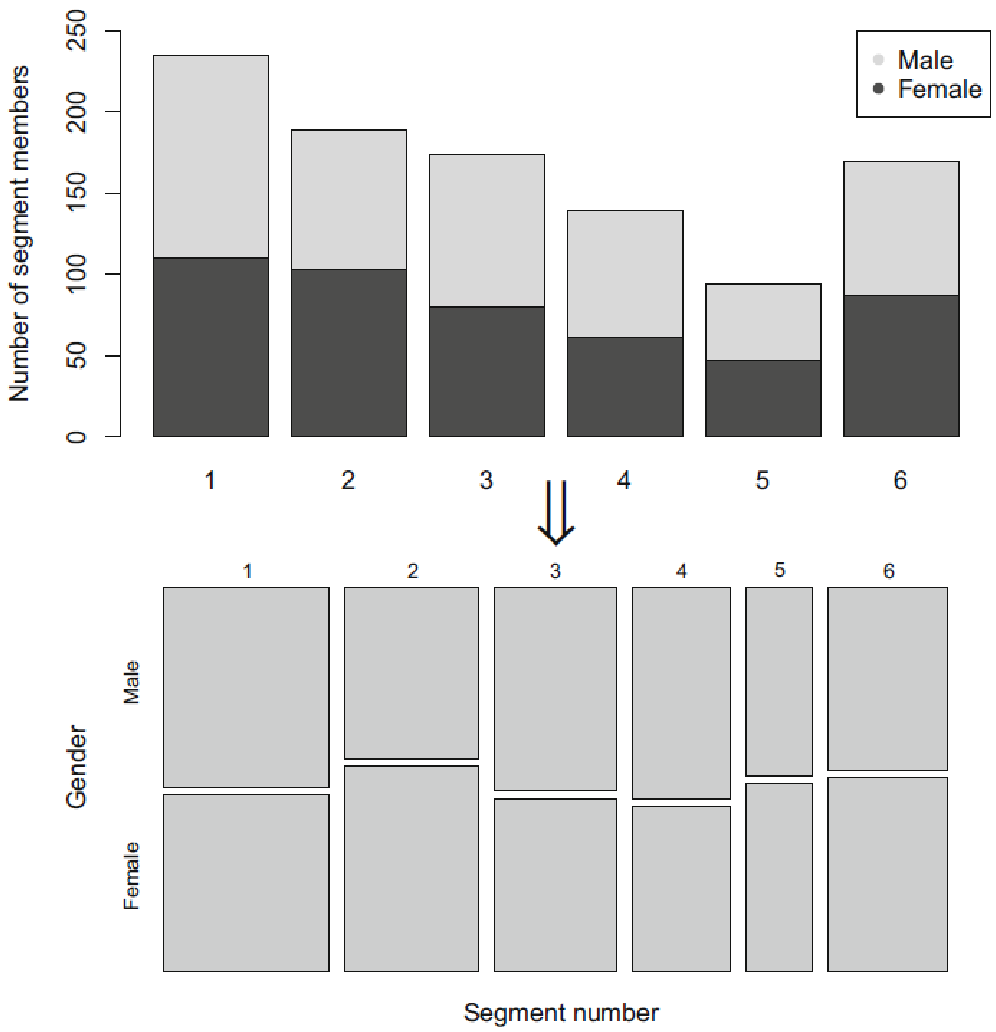
1	2	3	4	5	6
235	189	174	139	94	169

The easiest approach to generating a cross-tabulation is to add segment membership as a categorical variable to the data frame of descriptor variables. We use gender as the categorical variable in this case:

	Gender	
Segment Number	Male	Female
1	125	110
2	86	103
3	94	80

4	78	61
5	47	47
6	82	87

The text discusses the use of visualizations to examine gender differences across market segments. A stacked bar chart is presented in the upper panel of Figure given below, showing segment sizes and the distribution of males and females within each segment. However, comparing proportions across segments becomes challenging when segment sizes are unequal. An alternative solution is to use a mosaic plot, which represents absolute segment sizes as the width of the bars and the proportion of men and women as the height of rectangles within each segment. The advantage of the mosaic plot is that it allows for a direct comparison of proportions while still conveying the relative sizes of the market segments.



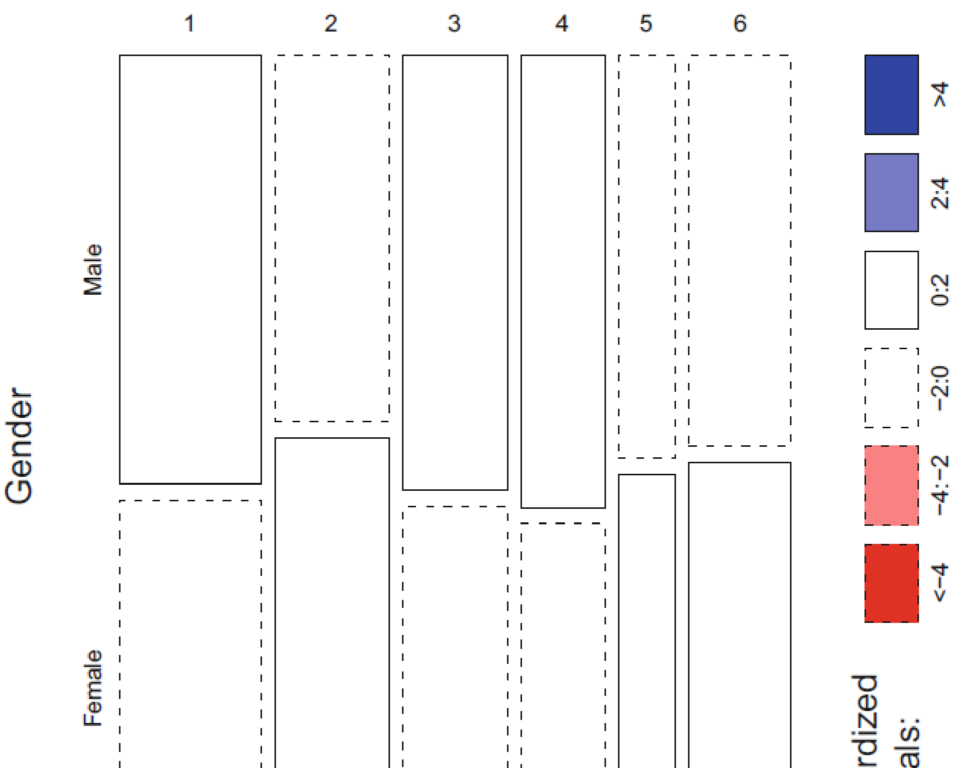
Mosaic plots are versatile visualizations that can also handle tables with multiple descriptor variables and incorporate elements of inferential statistics. They aid in interpretation by using colors to highlight discrepancies between observed and expected frequencies. The cell colors are based on the standardized difference between the expected and observed frequencies.

Negative differences are represented in red, indicating that the observed frequencies are lower than expected, while positive differences are shown in blue, indicating higher observed frequencies. The saturation of the color reflects the magnitude of the standardized difference. Standardized differences follow a standard normal distribution, with values typically falling within the range of [-2, 2] with a probability of approximately 95% and [-4, 4] with a probability of approximately 99.99%.

These standardized differences are equivalent to the standardized Pearson residuals from a log-linear model that assumes independence between the variables. Overall, mosaic plots provide a powerful tool for visualizing complex tables and examining the statistical significance of relationships between variables.

In the graph described below, different cell colors are used to represent contributions or standardized Pearson residuals. Dark red indicates contributions smaller than -4, light red represents contributions smaller than -2, white is used for values between -2 and 2 (considered not interesting), light blue indicates contributions larger than 2, and dark blue represents contributions larger than 4. The figure given below illustrates this color scheme with a corresponding legend.

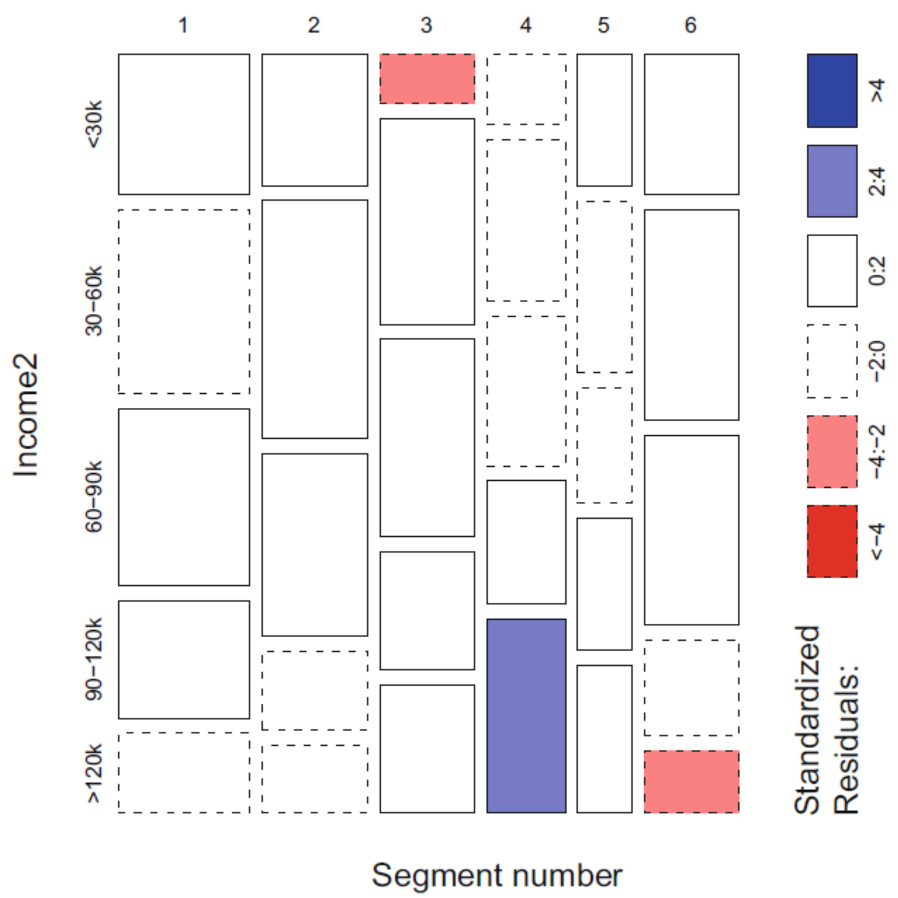
In the specific example shown in the figure given below, all cells are white, indicating that the six market segments derived from the Australian travel motives dataset do not exhibit significant differences in gender distribution. The proportion of male and female tourists is approximately equal across all segments. Additionally, dashed and solid borders around the rectangles indicate cells where the number of respondents is either lower or higher than expected, respectively.



]\The figure given below demonstrates a moderate association between segment membership and income. The top row corresponds to the lowest income category (less than AUD 30,000 per annum), while the bottom row represents the highest income category (more than AUD 120,000 per annum). The remaining three categories indicate income brackets of AUD 30,000 each, falling between these extremes.

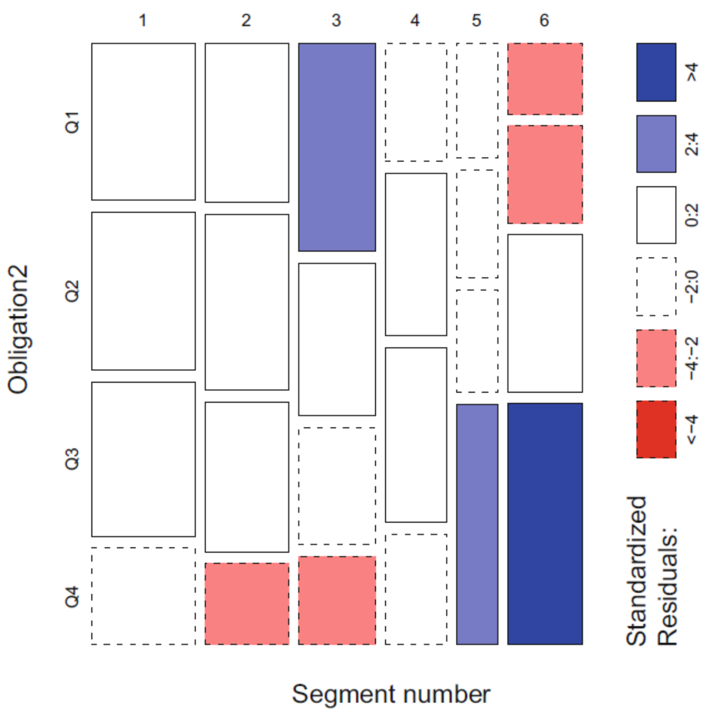
From the figure given below, we observe that individuals in segment 4 (column 4) - characterized by an interest in cultural offerings and local people - tend to have higher incomes. On the other hand, tourists with low income (top row) are less likely to be members of segment 3, which consists of individuals who prioritize luxury, fun, and entertainment during their vacations and seek to be pampered.

Furthermore, it is interesting to see that segment 6 (column 6) - the nature-loving segment - contains only a smaller proportion of individuals with very high incomes.



The figure given below illustrates a significant correlation between travel motives and the expressed moral obligation to protect the environment. The moral obligation score is derived by averaging responses to 30 survey questions that assess individuals' sense of responsibility towards engaging in environmentally friendly behaviors at home, such as not littering, recycling, and conserving water and energy. The scoring scale ranges from 1 (indicating the lowest moral obligation) to 5 (indicating the highest moral obligation), as respondents had five options to choose from. The sum of the scores ranges from 30 to 150, which is then rescaled to a range of 1 to 5 by dividing by 30.

To analyze this descriptor variable in its original metric format, please refer to the corresponding section provided below. In order to create the mosaic plot displayed in the figure given below, the moral obligation score is divided into four quarters, each representing 25% of the respondents. These quarters are labeled as Q1 (indicating low moral obligation) to Q4 (indicating high moral obligation). The variable "Obliged2" contains the recorded descriptor variable for this purpose.



Another interesting observation that can be made from the figure given above is how it visually represents the association between segment membership and the expressed moral obligation to protect the environment in a mosaic plot. In the figure given above we observe that segment 3 (column 3) - consisting of individuals seeking entertainment - has a significantly higher proportion of members with a low stated moral obligation to engage in environmentally friendly behaviors. Conversely, segment 3 has a significantly lower proportion of members in the high moral obligation category. On the other hand, segment 6 (plotted in column 6) displays the opposite pattern. Members of this segment, driven by a love for nature, exhibit a positive association with a high moral obligation to behave in an environmentally friendly way and a negative association with being in the lowest moral obligation category.

7.2.2 Metric Descriptor Variables

This portion will describe the use of the R packages "lattice" and "ggplot2" for creating conditional plots, which divide plots into sections or facets to display subsets of data, such as different market segments. The "lattice" package was used to generate a segment profile plot in a previous section.

In the context of segment description, the "lattice" package can be utilized to visualize the age distribution and the distribution of moral obligation scores for each segment. A new factor variable is created by combining the word "Segment" with the segment numbers, allowing segment names to be displayed in the plot. The histograms for age and moral obligation scores are generated using the "histogram()" function, with the "as.table" argument controlling the arrangement of panels.

The resulting histograms are shown in Figures 9.5 (age) and 9.6 (moral obligation). However, assessing the differences between market segments solely by looking at these plots can be challenging.

To gain further insights, a parallel box-and-whisker plot is utilized. This plot displays the distribution of the variable (e.g., age) separately for each segment. The "boxplot()" function is used to create the parallel box-and-whisker plot in R, specifying the variables and customizing the axis labels.

Overall, the text highlights the use of these R packages and provides examples of creating conditional plots, histograms, and parallel box-and-whisker plots to visualize and analyze differences between market segments.

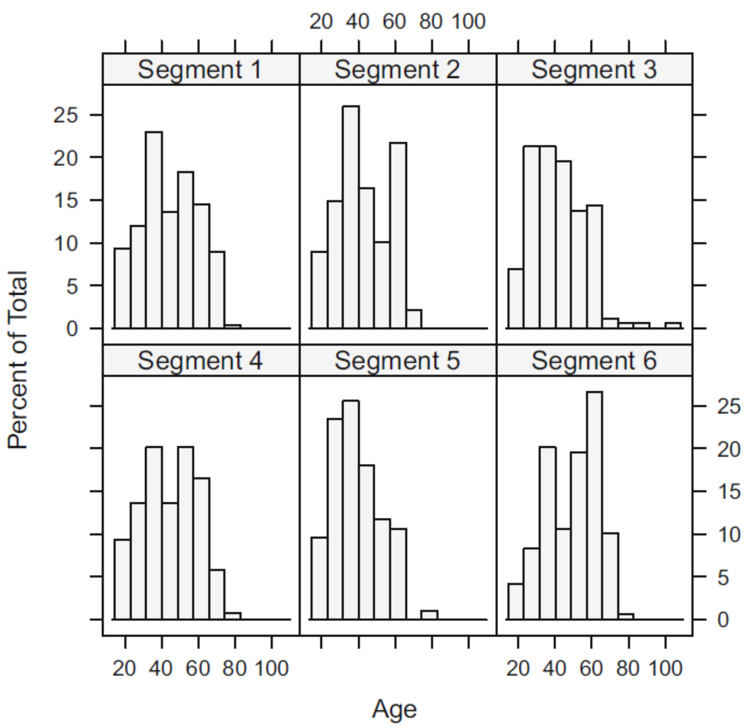


Fig. 9.5 Histograms of age by segment for the Australian travel motives data set

Now let’s discuss Figure 9.7, which shows a plot depicting differences in age across segments. The histograms examined earlier indicated that these differences are minimal. It can be observed that the median age of segment 5 members is slightly lower, while the median age of segment 6 members is slightly higher. However, it is important to subject these visual observations to statistical testing to confirm their significance.

Similar to mosaic plots, parallel box-and-whisker plots can incorporate elements of statistical hypothesis testing. In the provided R command, the width of the boxes can be made proportional to the size of market segments (varwidth = TRUE), and 95% confidence intervals for the medians can be included (notch = TRUE). The plot specifically focuses on the variable "Moral obligation" and the segment number as the x-axis label and "Moral obligation" as the y-axis label.

Overall, the text emphasizes the need to perform statistical testing to validate visually detected differences in descriptors and highlights how parallel box-and-whisker plots can

Fig. 9.8 Parallel box-and-whisker plot (with elements of statistical inference) of moral obligation to protect the environment by segment for the Australian travel motives data set

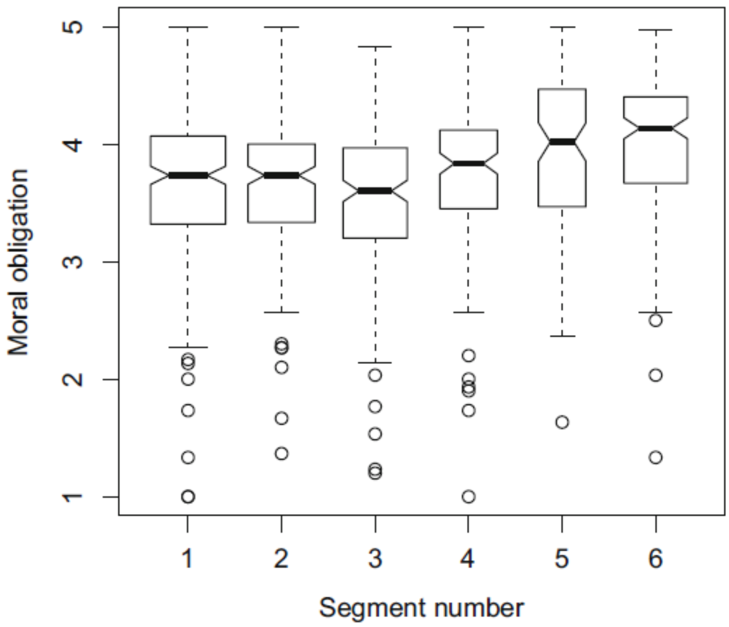
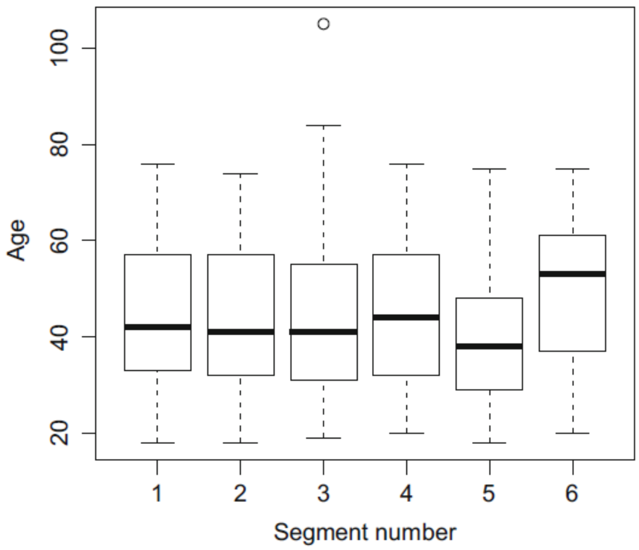


Fig. 9.7 Parallel box-and-whisker plot of age by segment for the Australian travel motives data set



be used for this purpose.
The resulting parallel box-and-whisker plot is displayed in Figure 9.8. This plot highlights several observations. Firstly, segment 5 is the smallest, indicated by its narrow box, while segment 1 is the largest. Additionally, members of segment 6 exhibit the highest moral obligation to protect the environment.

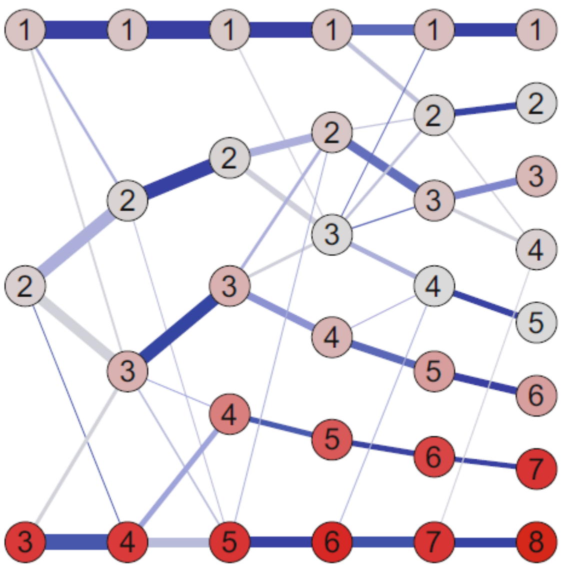
The notches in this version of the plot represent 95% confidence intervals for the medians. When the notches of different segments do not overlap, it usually indicates a significant difference. By inspecting Figure 9.8 alone, we can conclude that there is a significant difference in moral obligation between members of segment 3 and segment 6. The notches for these two segments are notably distant from each other. Although most of the boxes and whiskers are symmetric around the median, all segments contain some outliers at the lower end of moral obligation. This suggests that while most respondents claim to feel morally obliged to protect the environment, only a few openly admit to lacking a sense of moral obligation.

A modified version of the segment level stability across solutions (SLSA) plot can be employed to track the value of a metric descriptor variable across multiple market segmentation solutions. In this modified plot, different colors are used to represent the additional information contained in the metric descriptor variable for each node. The R command to generate this plot is as follows:

slsaplot(vacmot.k38, nodecol = vacmotdesc$Obligation)

Overall, these visualizations and analyses provide insights into the differences in moral obligation and demonstrate how the modified SLSA plot can incorporate metric descriptor variables for further exploration.

Fig. 9.9 Segment level stability across solutions (SLS_A) plot for the Australian travel motives data set for three to eight segments with nodes coloured by mean moral obligation values



The SLSA plot, displayed in Figure 9.9, represents each segment's average moral obligation to protect the environment using node colors. Deep red indicates high moral obligation, while light grey represents low moral obligation.

The bottom row of the plot consistently represents a segment that has been identified as potentially attractive (nature-loving tourists with an interest in the local population). This segment consistently exhibits high moral obligation across all analyzed segmentation solutions. Following this segment is another segment (segment 5 in the six-segment solution) that shows a tendency towards acquiescence bias, meaning respondents in this segment tend to agree with survey questions regardless of the content. Consequently, they are likely to express agreement when asked about their moral obligation to protect the environment.

In this modified SLSA plot, the shading of the edges represents the numeric SLSA value. Light grey edges indicate low stability values, while dark blue edges indicate high stability values.

Overall, the SLSA plot provides insights into the moral obligation of different segments, highlighting consistent patterns and potential biases within the data.

7.3 Testing for Segment Differences in Descriptor Variables:

Simple statistical tests can be used to assess differences in descriptor variables across market segments. For nominal or ordinal variables, such as gender or level of education, a chi-square test of independence can be performed. The p-value obtained from the test indicates the likelihood of observing the frequencies if there is no association between the variables. Significant p-values suggest differences between segments. However, if the p-value is not significant, as indicated by a chi-square test, no association is observed.

For associations between segment membership and moral obligation (a nominal variable), a chi-square test can also be applied. The mosaic plot provides a visual representation of the association, with color-coded cells indicating combinations that occur more or less frequently than expected under independence.

When dealing with metric variables like age or expenditure, differences in location (mean or median) across market segments can be assessed using parallel boxplots. Analysis of variance (ANOVA) is a commonly used method to test for significant differences in means among multiple groups. The F-test statistic and p-value from the ANOVA output indicate whether the mean values across segments are significantly different.

In addition to mean values, median values can be considered using the Kruskal-Wallis rank sum test as an alternative to ANOVA. This test compares medians among segments and provides p-values to determine if the medians are significantly different.

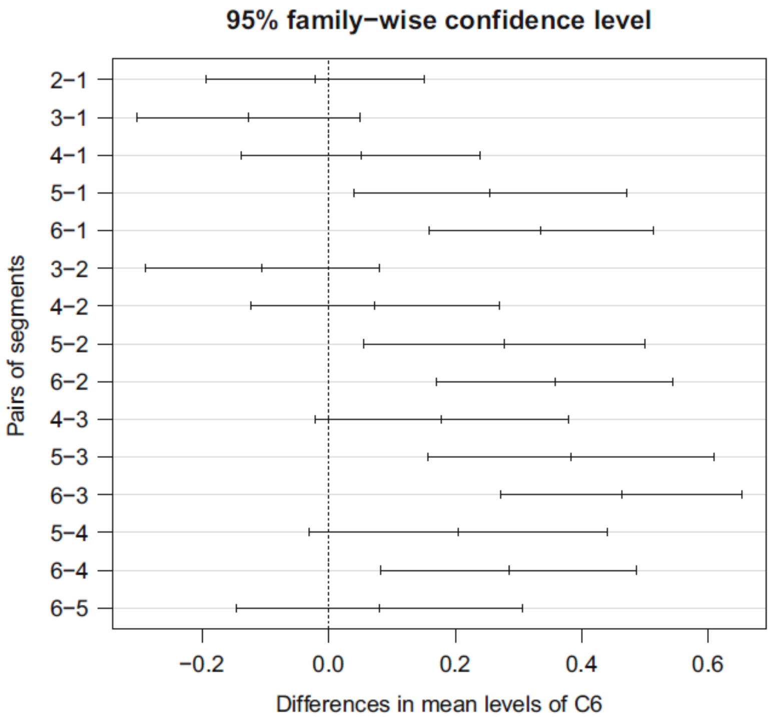
Overall, statistical tests and visualizations, such as chi-square tests, mosaic plots, ANOVA, and Kruskal-Wallis rank sum test, enable the assessment of differences in descriptor variables across market segments.

To identify which segments have significantly different mean levels of moral obligation, pairwise t-tests can be conducted. The output of the t-tests provides p-values that indicate whether the means of two segments are significantly different. The direction of the difference can be determined from a parallel box-and-whisker plot.

Adjusting p-values for multiple testing is essential when conducting a series of tests to assess a single hypothesis. Bonferroni correction is a conservative approach that multiplies all p-values by the number of tests. A less conservative method, proposed by Holm, can be used for adjusting p-values.

An alternative to pairwise t-tests is plotting Tukey's honest significant differences, which visually represents the comparisons of mean values between segments. Each row in the plot represents a pair of segments, with a horizontal solid line indicating the point estimate of the mean difference and the length of the line representing the confidence interval. If the confidence interval does not cross the vertical line at 0, the difference is considered significant.

In the case of moral obligation across segments, segments 1, 2, 3, and 4 do not differ significantly from each other. Similarly, segments 5 and 6 do not differ significantly from each other, but they exhibit significantly higher moral obligation compared to other segments, except for segments 4 and 5, which do not differ significantly. The parallel box-and-whisker plot supports these findings, showing the positioning of segment 4 between the low and high moral obligation groups.



7.4 Predicting Segments from Descriptor Variables

To gain insights into market segments, an alternative approach is to predict segment membership based on descriptor variables. This involves using a regression model where the categorical segment membership is the dependent variable, and the descriptor variables are the independent variables. These methods draw from statistical techniques for classification and machine learning for supervised learning. Unlike the methods discussed earlier, these approaches assess differences in all descriptor variables simultaneously. The prediction performance indicates how effectively market segments can be identified using the descriptor variables, and also reveals which variables are crucial for segment identification, particularly when variable selection methods are employed. Regression analysis serves as the foundation for these prediction models, assuming that the dependent variable y can be predicted by the independent variables x1, . . . , xp:

Regression models vary based on the function f(·), the assumed distribution of the dependent variable y, and the discrepancies between y and f(x1, . . . , xp).

The fundamental regression model is the linear regression model, which assumes that the function f(·) is linear and that y follows a normal distribution with a mean of f(x1, . . . , xp) and a variance of σ2. The relationship between the dependent variable y and the independent variables x1, . . . , xp can be expressed as follows:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p + \epsilon,$$

where ~ N(0, σ2).

In linear regression models, regression coefficients express how much the dependent variable changes if one independent variable changes while all other independent variables remain constant. The linear regression model assumes that changes caused by changes in one independent variable are independent of the absolute level of all independent variables.The dependent variable in the linear regression model follows a normal distribution.

7.4.1 Binary Logistic Regression

We can formulate a regression model for binary data using generalised linear models by assuming that f (y|μ) is the Bernoulli distribution with success probability μ, and by choosing the logit link that maps the success probability μ ∈ (0, 1) onto (−∞,∞) by

$$g(\mu) = \eta = \log\left(\frac{\mu}{1-\mu}\right).$$

The intercept in the linear regression model gives the mean value of the dependent variable if the independent variables x1, . . . , xp all have a value of 0. In binomial logistic regression, the intercept gives the value of the linear predictor η if the independent variables x1, . . . , xp all have a value of 0.

The other regression coefficients in a linear regression model indicate how much the mean value of the dependent variable changes if this independent variable changes while others remain unchanged. In binary logistic regression, the regression coefficients indicate how the linear predictor changes. The changes in the linear predictor corresponds to changes in the log odds of success. The odds of success are the ratio between the probability of success μ and the probability of failure 1−μ. If the odds are equal to 1, success and failure are equally likely. If the odds are larger than 1, success is more likely than failure. Odds are frequently also used in betting.

7.4.2 Multinomial Logistic Regression

Multinomial logistic regression can fit a model that predicts each segment simultaneously. Because segment extraction typically results in more than two market segments, the dependent variable y is not binary. Rather, it is categorical and assumed to follow a multinomial distribution with the logistic function as link Function.

7.4.3 Tree-Based Methods

Classification and regression trees (CARTs; Breiman et al. 1984) are an alternative modeling 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 visualizations, and the straight-forward incorporation of interaction effects. Classification and regression trees work well with a large number of independent variables. The disadvantage is that results are frequently unstable. Small changes in the data can lead to completely different trees.
The tree approach uses a stepwise procedure to fit the model. At each step, consumers are split into groups based on one independent variable. The aim of the split is for the resulting groups to be as pure as possible with respect to the dependent variable. This means that consumers in the resulting groups have similar values for the dependent variable. In the best case, all group members have the same value for a categorical dependent variable. Because of this stepwise splitting procedure, the classification and regression tree approach is also referred to as recursive partitioning.
The resulting tree (see Fig. 9.15) shows the nodes that emerge from each splitting step. The node containing all consumers is the root node. Nodes that are not split further are terminal nodes.We predict segment membership by moving down the tree. At each node, we move down the branch reflecting the consumer's independent variable. When we reach the terminal node, segment membership can be predicted based on the segment memberships of consumers contained in the terminal node.
Tree constructing algorithms differ with respect to:

- Splits into two or more groups at each node (binary vs. multi-way splits)
- Selection criterion for the independent variable for the next split
- Selection criterion for the split point of the independent variable
- Stopping criterion for the stepwise procedure
- Final prediction at the terminal node

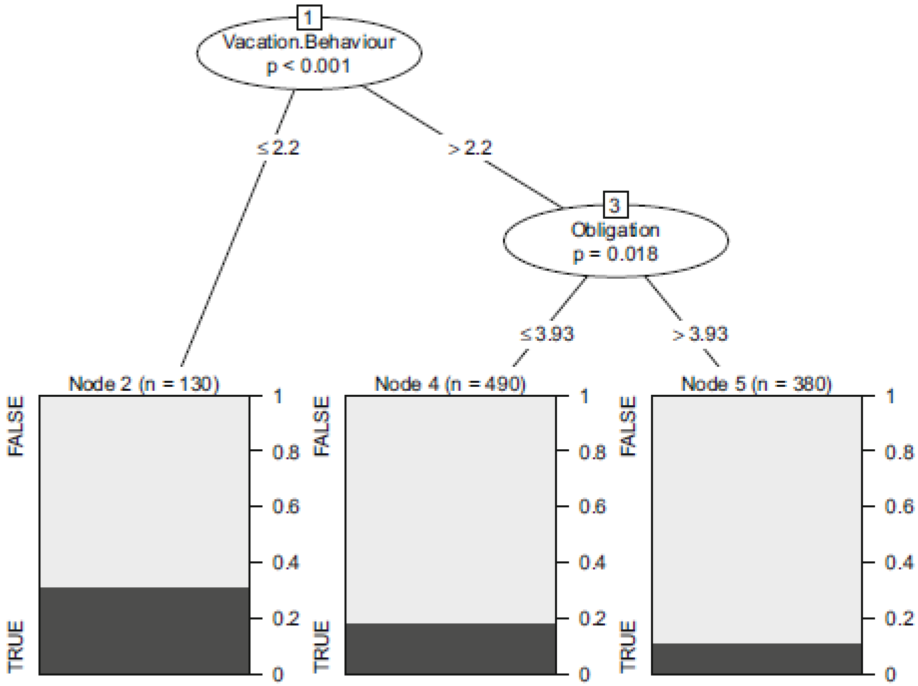


Fig. 9.15 Conditional inference tree using membership in segment 3 as dependent variable for the Australian travel motives data set

Step 8: Selecting the Target Segment(s) (Roshit Dahat)

8.1 The Targeting Decision

In Step 8 of market segmentation, the pivotal decision is made regarding which specific market segment(s) to target. This selection holds long-term significance as it profoundly impacts the organization's future performance. It signifies the transition from the exploration and evaluation phase to the commitment phase.

Once a global market segmentation solution has been chosen, typically in Step 5, there are several segments available for detailed examination. These segments are thoroughly profiled in Step 6 and described in Step 7. In Step 8, the objective is to choose one or more of these market segments for targeting. The segmentation team can draw upon the outcomes of Step 2, where knockout criteria for market segments were established, and segment attractiveness criteria were selected and weighted based on their relative importance to the organization.

Ideally, the knockout criteria should have already been applied in previous steps. For instance, during Step 6, market segments were profiled by assessing their key characteristics in relation to the segmentation variables. This evaluation would have revealed whether a segment is not sufficiently large, homogeneous, or distinct. Similarly, in Step 7, through a detailed description of the segments using descriptor variables, it would have become apparent whether a segment is not identifiable or reachable, or if it has needs that the organization cannot fulfill. For instance, imagine if the "BIG SPENDING CITY TOURIST" segment emerged as a highly attractive and distinct segment from the market segmentation analysis, but the destination conducting the analysis is a nature-based location in the outback of Australia. In this case, the chances of the destination meeting the needs of the highly attractive "BIG SPENDING CITY TOURIST" segment would be slim. Therefore, it is optimal for all the market segments under consideration in Step 8 to already meet the knockout criteria.

Nevertheless, it is always prudent to double-check. Hence, the initial task in Step 8 is to ensure that all the market segments still being considered as potential target markets have truly passed the knockout criteria test.

Once this verification is completed, the next step is to evaluate the attractiveness of the remaining segments and the organization's relative competitiveness in serving them. This evaluation involves addressing two key categories of questions:

1. Which market segment(s) does the organization most prefer to target and commit to?
2. Among the organizations offering the same product/service, which one would each segment most prefer to buy from? How likely is it that our organization would be chosen and committed to by each segment?

<u>The answers to these questions</u> form the foundation of the target segment decision, guiding the organization towards the <u>most suitable market segment(s) to focus on</u>.

10.2 Market Segment Evaluation

In target market selection, decision matrices are commonly used to assess the relative attractiveness of market segments and the organization's competitiveness in each segment. Various versions of decision matrices, such as the Boston matrix, General Electric/McKinsey matrix, and directional policy matrix, provide visualizations to aid in the evaluation and selection of target segments. The market segmentation team chooses the most suitable matrix variation for decision making.

The decision matrix considers two dimensions: segment attractiveness and relative organizational competitiveness. Segment attractiveness represents how desirable a segment is to the organization, while organizational competitiveness reflects the organization's appeal to the segment. In a segment evaluation plot, segments are depicted as circles, with their size indicating additional criteria like turnover contribution or loyalty.

Segment attractiveness and organizational competitiveness are subjective measures, and the ideal target segment is defined in Step 2 of the segmentation analysis. Weighted criteria for segment attractiveness are determined, considering the impact of each criterion on the overall attractiveness value. In Step 8, the segment selection stage, the actual values for each criterion within each segment need to be assigned. These values are derived from segment grouping, profiling, and description conducted in Steps 6 and 7.

To compute the position of each segment in the evaluation plot, the weight of each attractiveness criterion is multiplied by the assigned value for that criterion in each segment. The resulting weighted values are summed to represent the overall attractiveness of the segment (x-axis). This process is illustrated in Table 10.1, where example calculations are provided for five segment attractiveness criteria and their assigned weights. The same procedure is followed for evaluating relative organizational competitiveness.

The criteria considered for organizational competitiveness include product attractiveness, price suitability, distribution channel availability, and segment awareness or brand image. These criteria help understand how consumers select among alternative offers in the market.

Table 10.1 Data underlying the segment evaluation plot

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

By utilizing decision matrices and evaluating segment attractiveness and organizational competitiveness, the organization can make informed decisions regarding target segment selection.

The calculation of the value for the axis labeled "How attractive are we to the segment?" follows the same process as the attractiveness value from the organizational perspective. Criteria are agreed upon, weighted, and each segment is rated. The values are then multiplied and summed up. The data for the segment evaluation plot in Figure 10.1 can be found in Table 10.1.

The size of the bubbles in the plot represents another criterion, such as profit potential. Profit combines segment size and spending, which is crucial when selecting target segments. However, different criteria can be used depending on the context. For instance, a non-profit organization using market segmentation to recruit volunteers may use the number of hours volunteered as the bubble size.

With the completed plot, the segmentation team can engage in discussions. Based on Figure 10.1, segments 3 and 7 may be eliminated from consideration as they are less attractive compared to other segments, despite having high profit potential. Segment 5 is highly attractive but lacks affinity towards the organization. It is unlikely that the organization can successfully cater to this segment. Segment 8 is an excellent match as it is highly attractive to the organization and views the organization's offer favorably, although the profit potential is not very high. Therefore, segment 2 may need to be considered, as it has a positive affinity towards the organization, decent profit potential, and is equally attractive as segments 1, 4, and 6 (which do not favor the organization's offer).

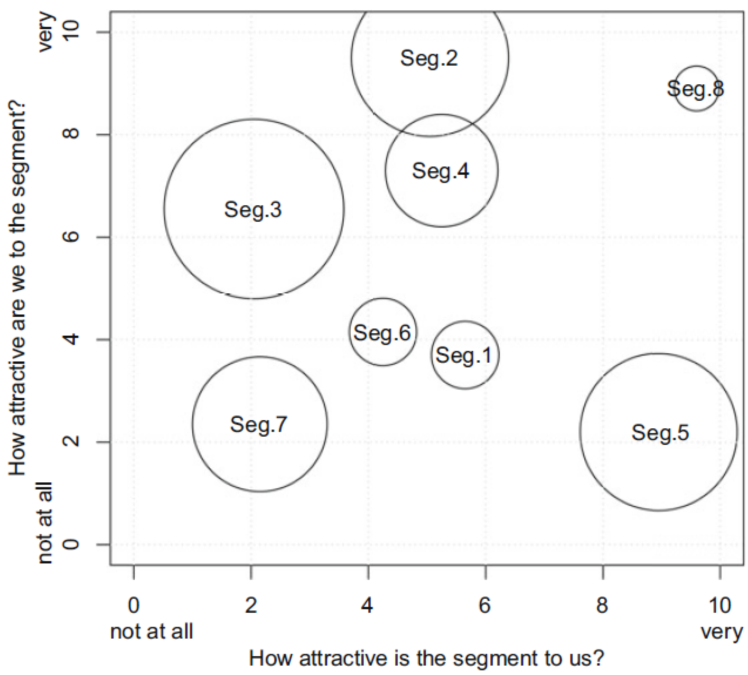


Fig. 10.1 Segment evaluation plot

Step 9 - Customising Market Mix (by Sumedh Deshkar)

To best ensure maximising the benefits of a market segmentation strategy, it is important to customise the marketing mix to the target segment. After the target segment(s) have been selected, then the characteristics of the target segments are mapped, to identify harnessable features. The marketing mix is understood to consist of the 4 Ps: Product, Price, Promotion and Place. Market segmentation does not stand independently as a marketing strategy. Rather, it goes hand in hand with the other areas of strategic marketing, most importantly: positioning and competition. In fact, the segmentation process is frequently seen as part of what is referred to as the segmentation-targeting-positioning (STP) approach. The segmentation-targeting-positioning approach postulates a sequential process. The process starts with market segmentation (the extraction, profiling and description of segments), followed by targeting (the assessment of segments and selection of a target segment), and finally positioning (the measures an organisation can take to ensure that their product is perceived as distinctly different from competing products, and in line with segment needs). The selection of one or more specific target segments may require the design of new, or the modification or re-branding of existing products (Product), changes to prices or discount structures (Price), the selection of suitable distribution channels (Place), and the development of new communication messages and promotion strategies that are attractive to the target segment (Promotion). One option available to the organisation is to structure the entire market segmentation analysis around one of the 4Ps. This affects the choice of segmentation variables. If, for example, the segmentation analysis is undertaken to inform pricing decisions, price sensitivity, deal proneness, and price sensitivity represent suitable segmentation variables

9.1 Product

One of the key decisions an organisation needs to make when developing the product dimension of the marketing mix, is to specify the product in view of customer needs. Often this does not imply designing an entirely new product, but rather modifying an existing one. Other marketing mix decisions that fall under the product dimension are: naming the product, packaging it, offering or not offering warranties, and after sales support services.

9.2 Price

Typical decisions an organisation needs to make when developing the price dimension of the marketing mix include setting the price for a product, and deciding on discounts to be offered.

9.3 Place

The key decision relating to the place dimension of the marketing mix is how to distribute the product to the customers. This includes answering questions such as: should the product be made available for purchase online or offline only or both; should the manufacturer sell directly to customers; or should a wholesaler or a retailer or both be used.

9.4 Promotion

Typical promotion decisions that need to be made when designing a marketing mix include: developing an advertising message that will resonate with the target market, and identifying the most effective way of communicating this message. Other tools in the promotion category of the marketing mix include public relations, personal selling, and sponsorship.

Step 10 - Evaluation and Monitoring (By Sumedh Deshkar)

After segmentation has been performed, two tasks remain, evaluating the effectiveness of the

strategy and continuous monitoring of the segments. Success of a strategy is evaluated usually by observing a target performance criterion such as sales, or amount of donations etc before and after use of the market segmentation strategy.
It is important to understand that the evaluated market segment analysis is like a picture of a moving object taken at a fixed moment in its motion. It thus becomes important to track the changes in the segments and their constituents. Over a period of time, a segment may remain unchanged, which is highly unlikely, or it may increase or decrease in size. New segments may arise, existing segments may disappear, multiple segments may merge to form a single segment or a single segment may split to form multiple segments. Changes in segment membership are problematic if (1) segment sizes change (especially if the target segment shrinks), and if (2) the nature of segments changes in terms of either segmentation or descriptor variables. Changes in segment size may require a fundamental rethinking of the segmentation strategy. Changes in segment characteristics could be addressed through a modification of the marketing mix. Monitoring thus is essential, and should be automated if possible, to obtain longitudinal data. This would help organisations be on top of any changes occurring within the market, giving them visible advantage over other players

Github Links for Python Code: **Sumedh Deshkar** **|** **Roshit Dahat** **|** **Shagun Shirohi**