# Market Segmentation Analysis


## Part I Introduction

### 1 Market Segmentation:

Market segmentation offers a multitude of advantages that empower organizations to excel in their marketing efforts:

1. Deeper Understanding: By segmenting the market, businesses gain valuable insights into consumer differences, enabling them to tailor their products and services to meet specific needs effectively.
2. Competitive Advantage: Focusing on distinct target segments allows organizations to stand out in the market and offer unique value propositions that cater precisely to consumers' preferences.

3. Enhanced Returns on Investment: Concentrating resources on targeted segments enables companies to optimize their marketing efforts, leading to higher returns on investments and increased profitability.
4. Potential Market Dominance: A well-implemented segmentation strategy can open the path to market dominance, securing a prosperous future for the organization and elevating its brand position.


# 2 Market Segmentation Analysis:

The Layers of Market Segmentation Analysis Market segmentation analysis involves three key layers of tasks:

1. Core Technical Tasks: This layer encompasses statistical processes that group consumers into segments based on product preferences or characteristics, providing a foundation for segmentation.
2. Additional Technical Tasks: Ensuring high-quality segmentation through comprehensive data collection, exploration, and segment profiling, which contributes to accurate and meaningful segment categorization.
3. Non-Technical Tasks: This layer focuses on addressing organizational implementation issues, such as decision-making processes and target segment selection, crucial for successful market segmentation.

Approaches to Market Segmentation Analysis Market segmentation analysis can be approached through two main methods:

1. Organizational Constraint-Based Approaches: Utilizing quantitative surveys, existing consumer classifications, or qualitative research to create segments based on predefined constraints and parameters.
2. Nature-Based Approaches: Segmenting based on single variables (e.g., demographics) or multiple variables (data-driven clustering) to identify naturally existing or artificially created market segments.

# Part II: Ten Steps of Market Segmentation Analysis

**Step 1 - Deciding (not) to Segment**

Implications of Committing to Market Segmentation Before embarking on market segmentation analysis, organizations must consider the implications:

1. Long-Term Commitment: Successful market segmentation requires dedication and substantial investments, demanding a long-term perspective to reap its benefits fully.
2. Costs: Implementation involves expenses related to research, surveys, focus groups, product modifications, and tailored communications. Organizations must ensure that the expected increase in sales justifies these costs.
3. Organizational Changes: Pursuing market segmentation may necessitate new product development, pricing adjustments, and tailored communications for different market segments, potentially impacting the internal structure of the organization.

Implementation Barriers Potential barriers that may hinder successful market segmentation implementation include:

1. Lack of Senior Management Involvement: Active commitment from top-level executives is vital for overcoming resistance and ensuring the successful adoption of segmentation strategies.

2. Organizational Culture: A culture not centered around consumer needs, resistance to change, and short-term thinking can hinder the successful implementation of market segmentation.
3. Lack of Training and Expertise: Adequate understanding of market segmentation concepts among senior management and the segmentation team is critical for effective execution.
4. Resource Constraints: Insufficient financial resources, lack of qualified personnel, or an inability to make necessary changes can pose significant obstacles.
5. Process-Related Barriers: Unclear objectives, lack of planning, and absence of structured processes can impede the development of the most suitable segmentation outcome.

Checklist to evaluate their readiness for market segmentation, organizations should answer the following questions:

1. Market Orientation: Is the organization customer-focused and dedicated to understanding consumer needs and preferences?
2. Willingness to Change: Is the organization open to change and committed to long-term segmentation strategies?
3. Openness to New Ideas: Is the organization receptive to new concepts and approaches that can enhance market understanding ?
4. Effective Communication: Is communication across organizational units effective to facilitate seamless implementation of market segmentation?
5. Capability for Structural Changes: Is the organization capable of making significant changes to align with segment-focused strategies, if required?
6. Sufficient Financial Resources: Does the organization possess adequate resources to support a successful segmentation strategy?

### Step 2 - Selecting Target Segments

Combining Attractiveness and Organizational Competitiveness Step 4 involves combining segment attractiveness and organizational competitiveness to identify the most suitable target segments. The segment evaluation plot, which assesses segment attractiveness against organizational competitiveness, is a valuable tool for this purpose.

1. The Segment Evaluation Plot

   The segment evaluation plot is a visual representation of segment attractiveness on one axis and organizational competitiveness on the other. This plotting helps the segmentation team visualize and compare potential target segments, aiding in decision-making.

2. Making Informed Decisions

   Analyzing the segment evaluation plot allows the team to prioritize segments that align with the organization's strengths and offer high market potential. The goal is to select target segments that offer the best balance between attractiveness and the organization's ability to cater to their needs.

3. Developing Marketing Strategies

   Tailoring the Marketing Mix After identifying the target segments, Step 5 involves developing tailored marketing strategies for each segment. This includes customizing the marketing mix, incorporating product offerings, pricing, distribution channels, and promotional activities that align with the specific needs and preferences of each segment.

4. Effective Communication

   Effective communication is crucial in successful market segmentation. Tailored messages and communication channels must be chosen to effectively reach and engage each target segment. Personalization and relevance in communication enhance customer engagement and brand loyalty.

5. Monitoring and Adaptation

   Market segmentation is an ongoing process that requires continuous monitoring of market dynamics, consumer behavior, and competitor activities. Regular assessment of chosen strategies and adaptation to changing market conditions ensure long-term success.

6. Implementing and Evaluating the Segmentation Strategy

   Implementation planning involves detailing the actions required to execute the chosen segmentation strategy. Specific timelines, responsibilities, and resource allocation must be defined to ensure a smooth implementation process.

7. Performance Measurement

   Effective performance measurement is essential to evaluate the success of the segmentation strategy. Key performance indicators (KPIs) relevant to each target segment should be established to track progress and identify areas for improvement.

8. Continuous Improvement

   The segmentation strategy should be viewed as a dynamic process, open to refinement based on performance insights. Continuous improvement allows organizations to stay relevant and responsive to changing market demands.

### Step 3 - Collecting Data for Market Segmentation

1. Segmentation Variables
   Market segmentation involves dividing a heterogeneous market into smaller, more homogenous groups based on specific characteristics known as segmentation variables. These variables are used to split the sample into distinct segments. Common segmentation variables include demographics, psychographics, behavior, and geographic location.

2. Segmentation Criteria
   In market segmentation, organizations must choose a segmentation criterion, which defines the type of information used for dividing the market. The common segmentation criteria include geographic, socio-demographic, psychographic, and behavioral factors. Bock and Uncles (2002) identify relevant consumer differences, such as profitability, bargaining power, benefit preferences, barriers to choice, and consumer interaction effects, as crucial considerations for segmentation.

   Selecting the appropriate criterion requires market knowledge and cannot be easily outsourced. There are no strict guidelines for the best criterion, but experts recommend using the simplest approach that effectively meets the product or service's needs at the least cost. Demographic or geographic segmentation might be sufficient, and choosing psychographic segmentation merely for its appeal can be unnecessary. The key is to use what works best for the specific marketing context and product or service.

   2.1 Geographic Segmentation:

   - Geographic information is the original and simplest segmentation criterion.
   - It involves using consumers' location of residence to form market segments.
   - This approach is practical for targeting customers from different regions with specific communication messages and channels.

- However, location alone does not necessarily reflect other relevant consumer characteristics or product preferences.

2.2 Socio-Demographic Segmentation:

- Socio-demographic criteria include age, gender, income, and education.
- Useful in industries like luxury goods, cosmetics, baby products, retirement villages, and tourism resorts.
- Provides clear segment membership for each consumer, but demographics alone may not explain all product preferences.
- Values, tastes, and preferences are considered more influential in consumers' buying decisions than socio-demographics.

2.3 Psychographic Segmentation:

- Groups people based on psychological criteria such as beliefs, interests, aspirations, and benefits sought.
- Benefit segmentation and lifestyle segmentation are popular psychographic approaches.
- Reflects underlying reasons for consumer behavior and preferences, but requires multiple segmentation variables due to its complexity.
- Depends on the reliability and validity of empirical measures capturing psychographic dimensions.

2.4 Behavioural Segmentation:

- Focuses on similarities in actual behavior or reported behavior.
- Uses behaviors like prior product experience, purchase frequency, amount spent, and information search behavior for segment extraction.
- Advantageous when based on actual behavior, as it directly groups people by their most relevant similarity.
- Avoids the need for developing valid measures for psychological constructs.
- Availability of behavioral data can be a challenge when including potential new customers in the segmentation analysis.

3. Data from Survey Studies

**3.1 Choice of Variables:**

- Carefully select segmentation variables for data-driven or commonsense segmentation.
- Include all relevant variables capturing the segmentation criterion while avoiding unnecessary ones.
- Unnecessary variables can lead to respondent fatigue and hinder algorithms from identifying the correct segmentation solution (noisy or masking variables).
- Conduct exploratory or qualitative research to develop a well-constructed questionnaire.

**3.2 Response Options:**

- Response options in surveys impact the scale of data for segmentation analysis.
- Binary or metric options are preferable for easier distance measures in data-driven segmentation.
- Ordinal data with unclear distances between options may complicate analysis.
- Visual analogue scales or binary options can be used when fine nuances need to be captured.

**3.3 Response Styles:**

- Survey data can be influenced by response biases, leading to response styles.
- Response styles can affect segmentation results, and algorithms may not differentiate between beliefs and response styles.
- Minimize the risk of capturing response styles during data collection to avoid misinterpretation of segments.

**3.4 Sample Size:**

- Sufficient sample size is crucial for accurate segmentation analysis.
- Sample size recommendations vary based on segmentation variables, number of segments, and market characteristics.
- A sample size of at least 60-70 times the number of segmentation variables is recommended for accurate segment recovery.
- Collect high-quality unbiased data with an adequate sample size to ensure reliable segmentation results.

4. Data from Internal Sources:

4.1 Internal data available to organizations, such as scanner data from grocery stores, booking data from airline loyalty programs, and online purchase data, represent actual consumer behavior.

4.2 The strength of such data lies in its accuracy and reliability as it reflects real consumer actions, avoiding biases and imperfections present in survey data.
4.3 Internal data is automatically generated and easily accessible if stored in a suitable format, requiring no extra effort for data collection.
4.4 However, internal data may be biased by over-representing existing customers, lacking information about potential future customers with different consumption patterns.

5. Data from Experimental Studies for Market Segmentation:

5.1 Experimental data, from field or laboratory experiments, can be used for market segmentation analysis.

5.2 Experimental data may involve testing responses to advertisements or conducting choice experiments and conjoint analyses.

5.3 Choice experiments and conjoint studies provide valuable information about consumer preferences for specific product attributes and levels, which can be used as segmentation criteria.

**Step 4: Exploring Data**

1. Data:

- Data exploration is a crucial step after data collection, involving cleaning and preprocessing the data, as well as identifying suitable algorithms for market segmentation.
- In the data exploration stage, several aspects are investigated, including the measurement levels of variables, the distribution of each variable individually, and the relationships between variables.
- Pre-processing and preparing the data for segmentation algorithms may be necessary to ensure accurate results.
- The mentioned travel motives data set contains information from 1000 Australian residents regarding their last vacation, with variables such as gender, age, education, income, occupation, and various travel motives.
- The summary of the data set provides insights into the characteristics of the variables, such as the number of respondents by gender, the age distribution (minimum, maximum, quartiles), and the presence of missing data (NAs).

2. Data Cleaning:

- Data cleaning is the first step before data analysis, involving checking for correct values and consistent labels for categorical variables.
- Plausible ranges for metric variables can be defined in advance to identify and correct any implausible values that may indicate errors in data collection or entry.
- Categorical variables should only contain permissible values, and any non-permissible values need to be corrected during the data cleaning process.
- In the Australian travel motives data set, no data cleaning is required for the variables Gender and Age. However, the variable Income2 needs to be re-ordered as the categories are not sorted in order.
- The re-ordering process involves copying the column to a helper variable, storing its levels, finding the correct re-ordering, and converting the variable into an ordered factor.
- Reproducibility is important in data cleaning, exploration, and analysis, as documenting and saving the steps taken allows for future replication and enables continuous monitoring and analysis of segmentation solutions.

3. Descriptive Analysis:

- Descriptive analysis: Descriptive analysis involves summarizing and understanding the data by using numerical and graphical representations. It helps in avoiding misinterpretation of complex analysis results.
- R's summary() function: In R, the summary() function provides a numeric summary of the data. For numeric variables, it returns the range, quartiles, and mean. For categorical variables, it returns frequency counts and the number of missing values.
- Graphical methods: Graphical methods, such as histograms, boxplots, scatter plots, bar plots, and mosaic plots, are useful for visualizing data. Histograms show the distribution of numeric variables, while bar plots display frequency counts of categorical variables. Mosaic plots illustrate associations between multiple categorical variables.
- Histograms in R: R offers various packages for constructing histograms. The example uses the "lattice" package. Histograms display the frequency of observations within specific value ranges (bins). By specifying the number of bins with the breaks argument, the plot can show finer details of the distribution.
- Boxplots in R: Boxplots summarize the distribution of numeric variables. R's boxplot() function generates a box-and-whisker plot, which includes the minimum, first quartile, median, third quartile, and maximum values. The plot helps identify skewness, outliers, and quartile ranges. Whiskers are typically limited in length to avoid emphasizing outliers.

4. Preprocessing

4.1 Categorical And Numerical Variables:

- Merging levels of categorical variables is useful when the original categories are too differentiated or have imbalanced frequencies. This process helps create more balanced categories that are easier to analyze.
- Categorical variables can be converted to numeric variables if the distances between adjacent scale points on the ordinal scale are approximately equal. This assumption is reasonable for variables like income, where categories cover ranges of equal length.
- The popular agreement scale, often referred to as the Likert scale, is another example of an ordinal scale that can be treated as numeric if the distances between answer options are assumed to be equal. However, response styles and cultural factors can affect the true distances, so careful consideration is required.
- Binary variables, such as dichotomous ordinal or nominal variables, can always be converted to numeric variables. Most statistical procedures work correctly with binary variables, and the conversion is straightforward by assigning 0 and 1 values.
- Numeric variables used for segmentation should be standardized to balance their influence. Standardization transforms variables to a common scale by subtracting the mean and dividing by the standard deviation. This ensures that variables with different ranges have comparable impact on segmentation results.
- Alternative standardization methods may be necessary when data contains outliers. Robust estimates, such as the median and interquartile range, can be used instead of the mean and standard deviation to minimize the influence of outliers.

4.2 Principal Component Analysis:

- PCA is a statistical technique used to transform a dataset with correlated variables into a new set of uncorrelated variables called principal components.
- The principal components are ordered by importance, with the first component capturing the most variability in the data, the second component capturing the second most, and so on.
- PCA retains the relative positions of observations (consumers) in the transformed dataset and does not change the dimensionality of the data.
- PCA can be applied to covariance or correlation matrices of numeric variables, and it is recommended to use the correlation matrix if the variables have different data ranges.
- In practice, PCA is often used to reduce the dimensionality of high-dimensional data for visualization purposes, typically by selecting the first few principal components that explain the most variation.

- The output of PCA includes the standard deviations of the principal components, which reflect their importance, and the rotation matrix, which specifies how the original variables contribute to each principal component.
- The summary function provides additional information about the PCA results, such as the proportion of variance explained by each principal component and the cumulative proportion of explained variance.
- The interpretation of the PCA results involves understanding which variables have the highest loadings (contributions) on each principal component. This helps identify the variables that are most important for each component.
- In the given example of the Australian travel motives dataset, the first principal component does not differentiate well between motives, while the second and third components display more distinct patterns.
- PCA can be useful for exploring data, identifying highly correlated variables, and potentially removing redundant variables from the segmentation base.

**Step 5 – Extracting Segments**

Segment Evaluation Criteria The third layer of market segmentation analysis involves incorporating user input into the evaluation process. Organizations define two sets of evaluation criteria for market segments:

- Knock-Out Criteria: These are essential features that segments must possess to be considered for targeting, ensuring alignment with organizational goals.
- Attractiveness Criteria: These criteria are used to evaluate the relative attractiveness of the remaining segments that meet the knock-out criteria.

Knock-Out Criteria Common knock-out criteria include:

- Homogeneity: Segments must consist of consumers who are similar to each other in specific characteristics.
- Distinctness: Each segment must have distinct characteristics that differentiate it from other segments.
- Size: The segment must be large enough to justify customizing the marketing mix.
- Match with Organizational Strengths: Segments should align with the organization's strengths to capitalize on its capabilities.
- Identifiability: Members of the segment must be identifiable in the marketplace.
- Reachability: There must be effective means of communicating with segment members.

Attractiveness Criteria Attractiveness criteria are used to rate segments based on their attractiveness, considering specific factors for each segment.

Implementing a Structured Process A structured process, including the use of segment evaluation plots, is essential for evaluating market segments' attractiveness and organizational competitiveness.

**Step 6 Profiling segments**

1. Identifying Key characteristics of Market segment

- Profiling is essential only when data-driven market segmentation analysis is used to understand the characteristics of resulting market segments
- For commonsense segmentation, profiling is not necessary as the segment characteristics are predefined based on selected variables (e.g., age groups).
- Profiling involves characterizing market segments individually and comparing them to other segments to identify defining characteristics.
- Good profiling is crucial for correct interpretation of segmentation results and making effective strategic marketing decisions.
- Graphical statistics approaches can make profiling less tedious and reduce the risk of misinterpretation.

2. Traditional Approaches to profiling market segments:

In this study, an Australian vacation dataset is used. Neural Gas clustering algorithm is used for extracting segments with parameters:- number of segments from 3 to 8 and 20 random restarts.

Neural Gas clustering algorithm:-

Neural gas is an artificial neural network, inspired by the self-organizing map. The neural gas is a simple algorithm for finding optimal data representations based on feature vectors. Neural gas has the advantage of robust convergence compared to online k-means clustering. It is mostly used in speech recognition and image processing for data compression or vector quantization.

```python
#import pandas and pickle library packages
import pandas as pd
import pickle

# Load the 'vacmot' dataset
vacmot = pd.read_csv('path/vacmot.csv')

# Load the saved clustering results from step-5 from the 'vacmot-clusters.pkl' file
with open('path/vacmot-clusters.pkl', 'rb') as file:
    vacmot_clusters = pickle.load(file)
```

Data-driven segmentation solutions are presented to users in two ways:

- High-level summaries that oversimplify segment characteristics and may be misleadingly trivial.
- Large tables providing exact percentages for each segmentation variable for each segment.

These tables are challenging to interpret and do not offer a quick overview of key insights. Below Table illustrates this issue, displaying mean values of segmentation variables by segment.

**Table 8.1** Six segments computed with the neural gas algorithm for the Australian travel motives data set. All numbers are percentages of people in the segment or in the total sample agreeing to the motives

| | Seg. 1 | Seg. 2 | Seg. 3 | Seg. 4 | Seg. 5 | Seg. 6 | Total |
|---|---|---|---|---|---|---|---|
| Rest and relax | 83 | 96 | 89 | 82 | 98 | 96 | 90 |
| Change of surroundings | 27 | 82 | 73 | 82 | 87 | 77 | 67 |
| Fun and entertainment | 7 | 71 | 81 | 60 | 95 | 37 | 53 |
| Free-and-easy-going | 12 | 65 | 58 | 45 | 87 | 75 | 52 |
| Not exceed planned budget | 23 | 100 | 2 | 49 | 84 | 73 | 51 |
| Life style of the local people | 9 | 29 | 30 | 90 | 75 | 80 | 46 |
| Good company | 14 | 59 | 40 | 58 | 77 | 55 | 46 |
| Excitement, a challenge | 9 | 17 | 39 | 57 | 76 | 36 | 33 |
| Maintain unspoilt surroundings | 9 | 10 | 16 | 7 | 67 | 95 | 30 |
| Cultural offers | 4 | 2 | 5 | 96 | 62 | 38 | 28 |
| Luxury / be spoilt | 19 | 24 | 39 | 13 | 89 | 6 | 28 |
| Unspoilt nature/natural landscape | 10 | 10 | 13 | 15 | 69 | 64 | 26 |
| Intense experience of nature | 6 | 8 | 9 | 21 | 50 | 58 | 22 |
| Cosiness/familiar atmosphere | 11 | 24 | 12 | 7 | 49 | 25 | 19 |
| Entertainment facilities | 5 | 25 | 30 | 14 | 53 | 6 | 19 |
| Not care about prices | 8 | 7 | 43 | 19 | 29 | 10 | 18 |
| Everything organised | 7 | 21 | 15 | 12 | 46 | 9 | 16 |
| Do sports | 8 | 12 | 13 | 10 | 46 | 7 | 14 |
| Health and beauty | 5 | 8 | 10 | 8 | 49 | 16 | 12 |
| Realise creativity | 2 | 2 | 3 | 8 | 29 | 14 | 8 |

- In the context of binary travel motives, segment means represent the percentage of members engaging in each activity.
- To identify defining characteristics of market segments, one must compare segment percentage values for each segmentation variable with values of other segments or the total value in the far-right column.

Example Insights from Table :

3. Defining Characteristics of Segment 2 are ;

- Being motivated by rest and relaxation, and not wanting to exceed the planned travel budget.
- Many members of segment 2 care about a change of surroundings, but not about cultural offers, an intense experience of nature, about not caring about prices, health and beauty and realising creativity.

Segment 1 is likely to be a response style segment because as for each travel motive ,the percentage of segment members indicating that a travel motive is relevant to them is low compared to the overall percentage of agreement.

4. Problem of Traditional approach (Comparison of segments):

Consider the above table. There are total 6 Segments and 20 Segment variables. Total number of Comparison required would be = 6 * 20 = 120 Comparisons.

Now assume each Segment value is to be compared with Other segment values.

Total number of pairs to be compared are = 6C2 = 6!/(2!(6-2)!) = 15 pairs

And these 15 pairs need to be compared for each of 20 Segment variables so total 300 comparisons.

So there will be a total 420 Comparison including those between segments only and between segments and the total. This would be very tedious task.

To deal with this -sometimes information is provided about statistical significance of difference between segments of each variable ,which is not statistically correct as Segments are derived directly from Segmentation variables and are maximally different so use of standard statistical test is not allowed.

5. Segment Profiling with Visualizations:

Market segmentation solutions often lack effective use of graphics in their presentation, relying on highly simplified or complex tabular representations, although Data visualization using graphics is an integral part of statistical data analysis and is particularly important in exploratory statistical analysis, such as cluster analysis, as it provides insights into complex relationships between variables.

Visualizations offer a simple way of monitoring developments over time, especially in the context of big data.

The use of visualizations in the data-driven market segmentation process helps in inspecting and understanding segments in detail, facilitating the interpretation of segment profiles, and aiding in the critical decision-making process of selecting the most suitable solution.

6. Identifying Defining Characteristics of Market Segments:

A good way to understand the defining characteristics of each segment is to produce a segment profile plot. The segment profile plot shows – for all segmentation variables – how each market segment differs from the overall sample

Here We will study two methods of Segment profiling using Visualization:

- Hierarchical plot
- Panel Plot

1. Hierarchical plot:

Hierarchical plot is generated by ordering segmentation variables by similarity of answer patterns. We can achieve this by clustering the columns of the data matrix:

```
# Import required libraries
import numpy as np
from scipy.spatial.distance import pdist   # calculates the distance matrix
from scipy.cluster.hierarchy import linkage   #H for hiarachical Clustering


vacmot_vdist = pdist(vacmot.T)  # Transposes vacmot dataframe
vacmot_vclust = linkage(vacmot_vdist, method='ward')
```

This code calculates the distance matrix vacmot_vdist using pdist and then performs hierarchical clustering using the "ward" method with linkage. Make sure to have the numpy and scipy libraries installed in your Python environment.

Generated plot :

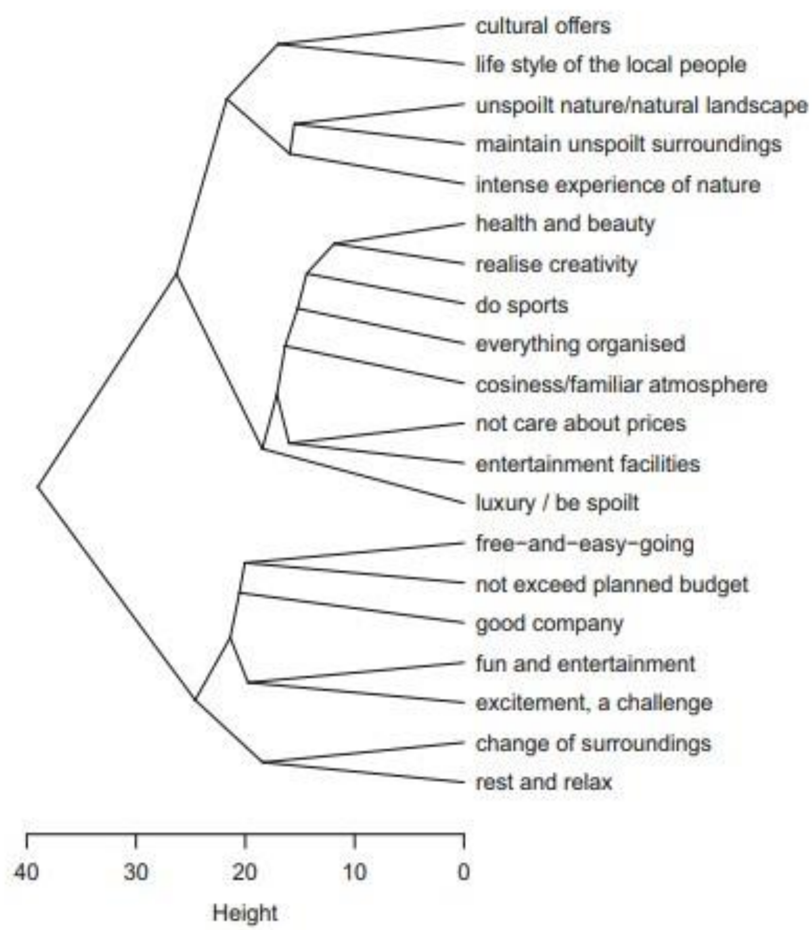


**Fig. 8.1** Hierarchical clustering of the segmentation variables of the Australian travel motives data set using Ward's method

Example of Segment profile :

Tourists who are motivated by cultural offers are also interested in the lifestyle of local people.

Tourists who care about an unspoilt natural landscape also show interest in maintaining unspoilt surroundings, and seek an intense experience of nature.

2. Panel plot :

Each panel represents one segment.

For each segment, the segment profile plot shows the cluster centres (centroids, representatives of the segments). These are the numbers contained in the Table shown in the Traditional approach section.

The dots in Fig represent the total mean values for the segmentation variables across all observations in the data set. The dots are the numbers in the last column in Table.

```python
# Get the order of segments from the hierarchical clustering
order = vacmot_vclust['leaves'][::-1]

# Sort the data based on the clustering order
sorted_vacmot = vacmot.iloc[:, order]

# Create the bar chart
plt.bar(range(len(vacmot_k6.labels_)), sorted_vacmot.mean(axis=1), color='colour', alpha=0.7)
plt.xlabel('Segments')
plt.ylabel('Mean Value')
plt.title('Segment Mean Values')
plt.show()
```

Result Panel plot :

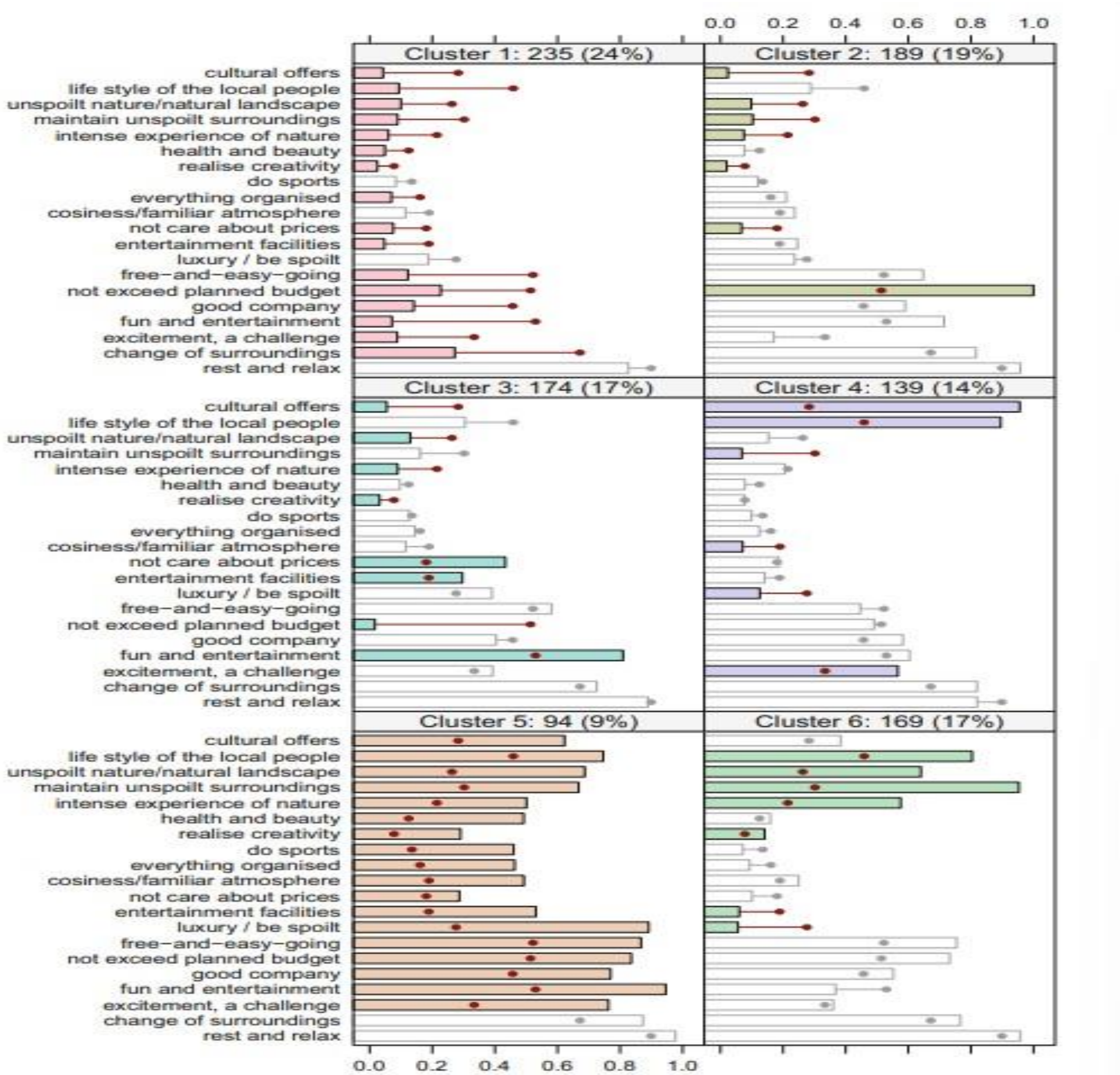


**Fig. 8.2** Segment profile plot for the six-segment solution of the Australian travel motives data set

Note :

- A variable with a segment mean significantly higher than the total sample mean is considered a marker variable.
- The absolute difference between the segment mean and the total sample mean should be substantial, often indicated by a deviation of 0.25 or a relative difference of 50% from the total mean.
- Marker variables are particularly useful for identifying characteristics that differentiate segments from the overall population.
- The example of the travel motive "HEALTH AND BEAUTY" in Figure illustrates that variables with low sample means (e.g., 0.12) can still be considered marker variables if they deviate significantly from the mean (e.g., by 50% of 0.12, which is 0.06).

Conclusion :-

- Using a segment profile plot for presenting segmentation results is faster and easier to interpret compared to using a table format, regardless of the table's structure.
- The plot helps identify marker variables and characteristics that differentiate segments.
- Strategic decisions based on segmentation analysis involve significant financial commitments, making good visualizations essential for an excellent return on investment.

## Assessing Segment Separation :

- Segment separation can be visualized using a segment separation plot which shows the overlap of segments across relevant dimensions of the data space.
- Segment separation plots are straightforward when the number of segmentation variables is low and become more complex with increasing number of Segment variables or dimensions.
- Despite complexity, segment separation plots provide a quick and useful overview of the data situation and segmentation solution.

Fig below provides examples of segment separation plots for **two different data sets**: one with **three well-separated segments** and the other with an **elliptic data structure.**

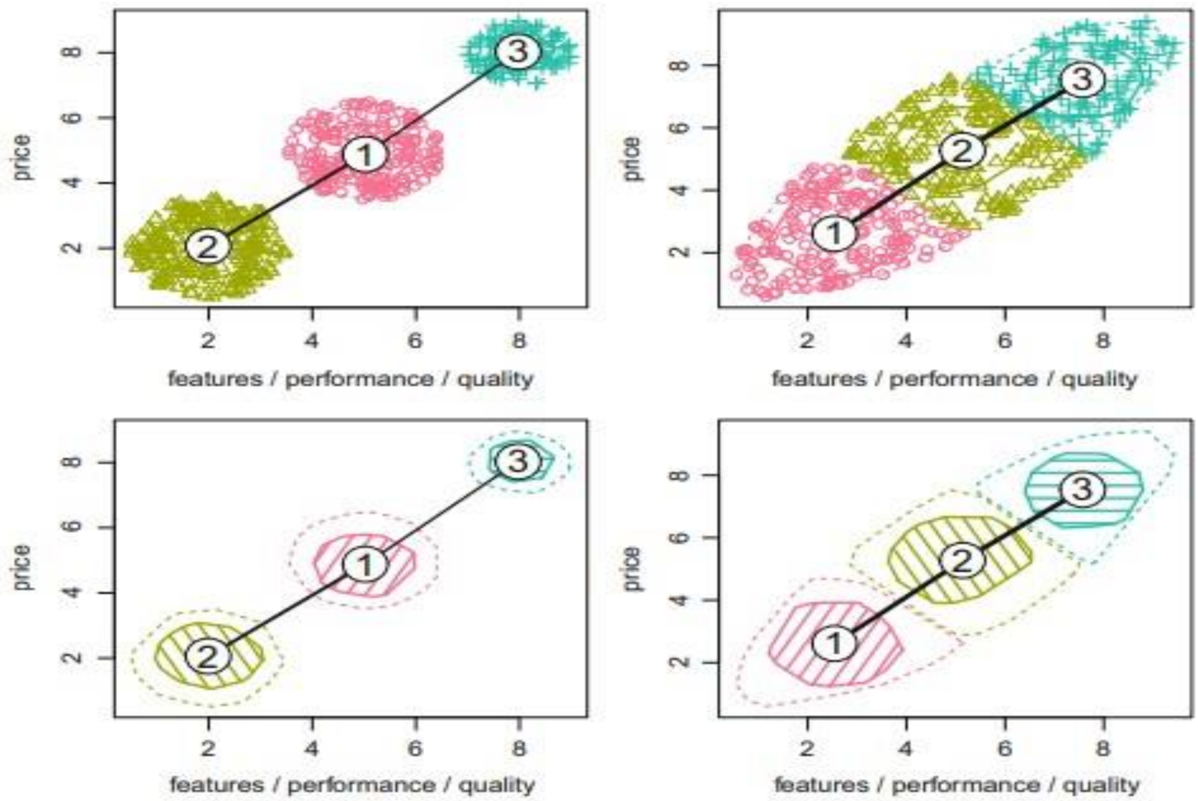


**Fig. 8.4** Segment separation plot including observations (first row) and not including observations (second row) for two artificial data sets: three natural, well-separated clusters (left column); one elliptic cluster (right column)

- The plots consist of a scatter plot of the observations colored by segment membership, along with cluster hulls and a neighborhood graph.
- The cluster hulls represent the shape and spread of the true segments, with dash hulls containing all observations and solid hulls containing approximately half of the observation.
- Neighborhood graphs indicate similarity between segments, The black lines connect segment centers, and indicate similarity between segments,with thicker lines representing more observations having two segment centers as their closest.

For a 20 dimensional Travel motive data set, such 2d plot is not possible. In such a situation, the 20- dimensional space needs to be projected onto a small number of dimensions to create a segment separation plot. This can be achieved by Principal component analysis.

```python
from sklearn.decomposition import PCA   #import PCA Class


pca = PCA()
vacmot_pca = pca.fit_transform(vacmot)   # Perform PCA on Vacmot dataset
```

The vacmot_pca variable will now contain the transformed data after applying PCA. The PCA class automatically centers the data before performing PCA.

This provides the rotation applied to the original data when creating our segment separation plot. We use the segmentation solution obtained from neural gas ,and create a segment separation plot for this solution:

```python
# module to plot the projection axes on the same plot.
from matplotlib.mlab import projAxes
# vacmot_k6 is a  pandas DataFrame containing the clustered data
# vacmot_pca is the result of PCA transformation on the vacmot data

# Extract the second and third principal components from the PCA result
pc2 = vacmot_pca[:, 1]
pc3 = vacmot_pca[:, 2]

# Create a scatter plot with the projected data
plt.scatter(pc2, pc3, c=vacmot_k6, cmap='rainbow')
plt.xlabel("Principal Component 2")
plt.ylabel("Principal Component 3")
projAxes(plt.gca(), 2, 3, linewidth=2, color='black')
plt.colorbar()
plt.show()
```

**Result plot** :

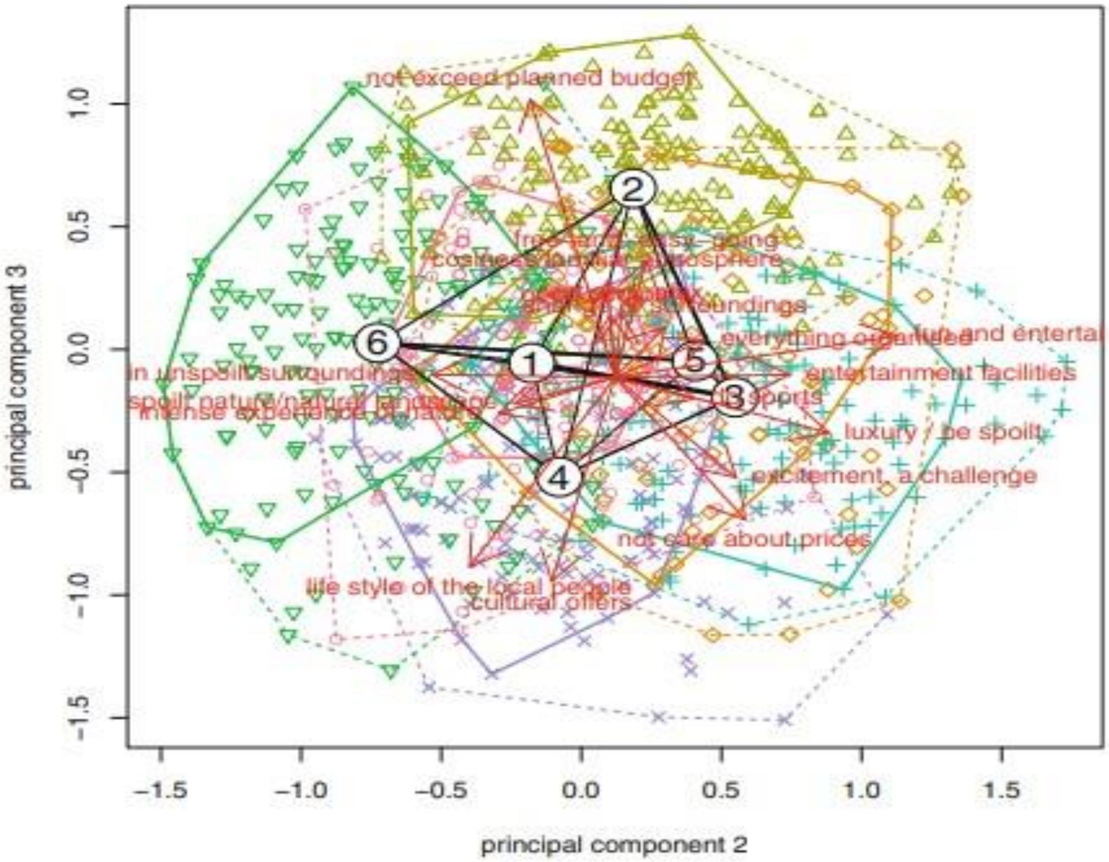


**Fig. 8.5** Segment separation plot using principal components 2 and 3 for the Australian travel motives data set

Due to the overlap of market segments the plot is messy and hard to read. Improvising visualization by modifying color, omitting observations and highlighting only the inner area of each segment to a cleaner version.

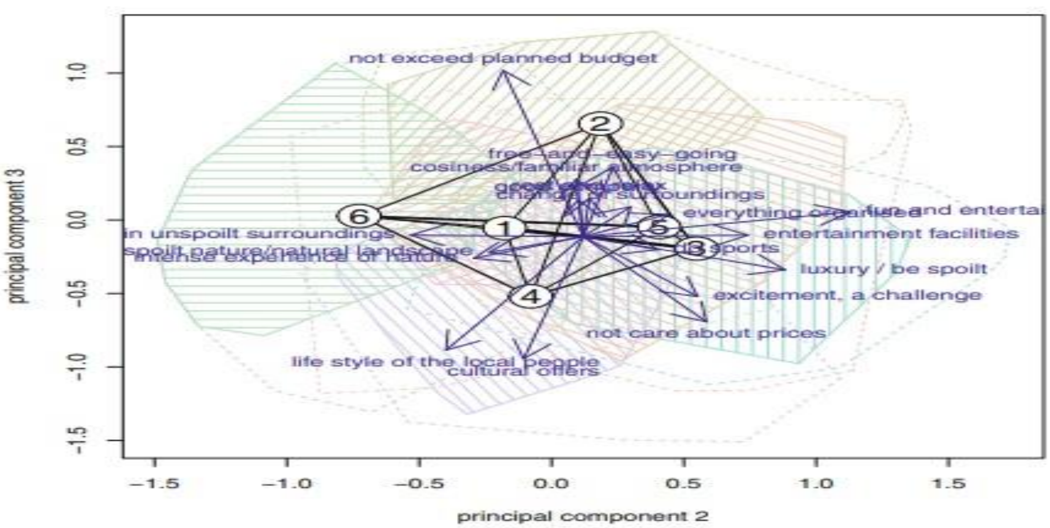


**Fig. 8.6** Segment separation plot using principal components 2 and 3 for the Australian travel motives data set without observations

Figure shows the existence of a market segment (segment 6, green shaded area) that cares about maintaining unspoilt surroundings, unspoilt nature, and wants to intensely experience nature when on vacations. Exactly opposite is segment 3 (cyan shaded area) wanting luxury, wanting to be spoiled, caring about fun, entertainment and the availability of entertainment facilities, and not caring about prices.

## Step 7: Describing segments

When describing differences between market segments in one single nominal or ordinal descriptor variable, the basis for all visualisations and statistical tests is a cross-tabulation of segment membership with the descriptor variable.

- Nominal and ordinal descriptor variables :
  When segments are unequal, it becomes difficult to compare the proportions of variable values. In such cases, mosaic plots can be used.
- Metric descriptor variables : Use of structures like histograms can be employed.
- We can use a modified version of the segment level stability across solutions (SLSA) plot to trace the value of a metric descriptor variable over a series of market segmentation solutions.

**Testing for Segment Differences in Descriptor Variables**

Statistical tests can be used to formally test for differences in descriptor variables across market segments.
The outcome of the segment extraction step is segment membership, the assignment of each consumer to one market segment. Segment membership can be treated like any other nominal variable. It represents a nominal summary statistic of the segmentation variables. Therefore, any test for association between a nominal variable and another variable is suitable.

The appropriate test for independence between columns and rows of a table is the $\chi2$-test.

- Points about $\chi2$-test :
  1. The p-value indicates how likely the observed frequencies occur if there is no association between the two variables.
  2. Small p-values (typically smaller than 0.05), are taken as statistical evidence of differences in the gender distribution between segments.

- The most popular method for testing for significant differences in the means of more than two groups is Analysis of Variance (ANOVA).
- The analysis of variance performs an F-test with the corresponding test statistic given as F value. The F value compares the weighted variance between market segment means with the variance within market segments. Small values support the null hypothesis that segment means are the same.
- As a robust alternative we can report median values by segment, and calculate p-values of the Kruskal-Wallis rank sum test. The Kruskal-Wallis rank sum test assumes (as null hypothesis) that all segments have the same median.

**Predicting Segments from Descriptor Variables**

- We use a regression model with the segment membership as categorical dependent variable, and descriptor variables as independent variables.

Regression analysis is the basis of prediction models. Regression analysis assumes that a dependent variable y can be predicted using independent variables or regressors $x1,\ldots, xp$:

$y \approx f(x1,\ldots,xp)$

- Regression models differ with respect to the function $f(\cdot)$, the distribution assumed for y, and the deviations between y and $f(x1,\ldots,xp)$.

The linear regression model assumes that function $f(\cdot)$ is linear, and that y follows a normal distribution with mean $f(x1,\ldots,xp)$ and variance $\sigma^2$.
The relationship between the dependent variable y and the independent variables $x1,\ldots,xp$ is given by:
$y = \beta0 + \beta1x1 + \ldots + \beta pxp +$ , where $\sim N(0, \sigma^2)$.

- 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.
-

In the linear regression model, the mean value of y given $x1,\ldots,xp$ is modelled by the linear function: $E[y|x1,\ldots,xp] = \mu = \beta0 + \beta1x1 + \ldots + \beta pxp$

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. Generalised linear models (Nelder and Wedderburn 1972) can accommodate a wider range of distributions for the dependent variable. This is important if the dependent variable is categorical, and the normal distribution, therefore, is not suitable.

Generalised linear models y are not limited to the normal distribution. We could, for example, use the Bernoulli distribution with y taking values 0 or 1. In this case, the mean value of y can only take values in (0, 1). It is therefore not possible to escribe the mean value with a linear function which can take any real value. Generalised linear models account for this by introducing a link function $g(\cdot)$. The link function transforms the mean value of y given by $\mu$ to an unlimited range indicated by $\eta$. This transformed value can then be modelled with a linear function: $g(\mu) = \eta = \beta 0 + \beta 1x1 + ... + \beta pxp$. $\eta$ is referred to as a linear predictor. We can use the normal, Poisson, binomial, and multinomial distribution for the dependent variable in generalised linear models. The binomial or multinomial distribution are necessary for classification. A generalised linear model is characterised by the distribution of the dependent variable, and the link function. In the following sections we discuss two special cases of generalised linear models: binary and multinomial logistic regression. In these models the dependent variable follows either a binary or a multinomial distribution, and the link function is the logit function.

**Binary Logistic Regression :**

- We can formulate a regression model for binary data using generalised linear models by assuming that $f(y|\mu)$ is the Bernoulli distribution with success probability $\mu$, and by choosing the logit link that maps the success probability $\mu \in (0, 1)$ onto $(-\infty,\infty)$ by $g(\mu) = \eta = \log( \mu \ / 1 - \mu \ )$

- 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 $\eta$ if the independent variables x1,...,xp all have a value of 0. The probability of being in segment 3 for a respondent with age 0 and a low moral obligation value is calculated by transforming the intercept with the inverse link function, in this case the inverse logit function: $g{-}1(\eta) = \exp(\eta) / 1 + \exp(\eta)$.

Transforming the intercept value of −0.72 with the inverse logit link gives a predicted probability of 33% that a consumer of age 0 with low moral obligation is in segment 3.

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 correspond to changes in the log odds of success. The odds of success are the ratio between the probability of success $\mu$ and the probability of failure $1-\mu$. 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.

**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.

- With function Anova() we assess if dropping a single variable significantly reduces model fit. Dropping a variable corresponds to setting all regression coefficients of this variable to 0. This means that the regression coefficients in one or several columns of the regression coefficient matrix corresponding to this variable are set to 0.

**Tree-Based Methods :**

Classification and regression trees (CARTs; Breiman et al. 1984) are an alternative modelling approach for predicting a binary or categorical dependent variable given a set of independent variables. Classification and regression trees are a supervised learning technique from machine learning. The advantages of classification and regression trees are their ability to perform variable selection, ease of interpretation supported by visualisations, and the straight-forward incorporation of interaction effects. Classification and regression trees work well with 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.

- Recursive partioning :

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 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

**Step 8: Selecting the Target Segment(s):**

1. The Targeting Decision:
   - Step 8 involves selecting one or more market segments for targeting, which significantly impacts the organization's future performance.
   - After a global market segmentation solution is chosen, several segments are available for inspection and profiling.
   - The segments should have already met the knock-out criteria established in Step 2, ensuring they are large, homogeneous, identifiable, and meet the organization's capabilities.

2. Market Segment Evaluation:
   - Target market selection is often visualized using a decision matrix to assess segment attractiveness and organizational competitiveness.
   - Various decision matrices are available, such as the Boston matrix, General Electric/McKinsey matrix, etc.
   - The two criteria plotted along the axes are segment attractiveness and relative organizational competitiveness.
   - The attractiveness of each segment is determined by applying weights to specific attractiveness criteria from Step 2 and combining them based on the segment's profiles and descriptions.
   - Similarly, the relative organizational competitiveness of each segment is assessed using criteria like product appeal, price suitability, distribution channels, and segment awareness.

3. Segment Evaluation Plot:
   - A generic segment evaluation plot can be used to visualize the relative attractiveness and competitiveness of each segment.
   - The x-axis represents "How attractive is the segment to us?" based on weighted values of segment attractiveness criteria.
   - The y-axis represents "How attractive are we to the segment?" based on weighted values of organizational competitiveness criteria.
   - Bubble sizes on the plot can represent other critical values, such as profit potential, loyalty, or volunteering hours.

4. Using the Segment Evaluation Plot:
   - The plot facilitates discussions within the segmentation team to compare and select target segments.

- Segments that perform poorly in attractiveness or competitiveness may be eliminated from consideration.
- Target segments may be chosen based on a combination of attractiveness, organizational fit, and profitability.

Note: The text provides guidance on how to create the segment evaluation plot using R and emphasizes the importance of considering the ideal target segment criteria specified in Step 2 during the target segment selection process.

**Step 9: Selecting the Target Segment(s):**

1. Implications for Marketing Mix Decisions
    - Marketing mix consists of various elements that contribute to achieving sales results.
    - The traditional marketing mix is commonly referred to as the 4Ps: Product, Price, Promotion, and Place.
    - Market segmentation is a crucial part of the segmentation-targeting-positioning (STP) approach in strategic marketing.
2. Product
    - Customizing the product to cater to the needs of the target segment is essential.
    - For example, a destination with a rich cultural heritage targeting a segment interested in museums and monuments might create a product package called "Museums, Monuments & Much, Much More."
3. Price
    - Pricing decisions should align with the segment's willingness to pay.
    - In the example, if the target segment (e.g., segment 3) spends more per person per day, the destination can consider pricing the product with a premium.
4. Place
    - Distribution decisions should consider how the product reaches the target segment.
    - For example, if the target segment prefers online booking, the destination should ensure an online booking option for their product.
5. Promotion
    - Promotional strategies should be designed to resonate with the target segment.
    - Information sources and media preferences of the target segment should be considered when advertising the product.
    - In the example, segment 3 relies heavily on information from tourist centers and prefers Channel 7, indicating targeted promotional efforts.

**CASE STUDY: Mcdonalds Fast Food**

**Github Repo Links:**

Pranav Chouhan: https://github.com/PranavChouhan10/Feynn_Labs/tree/main

Utkarsh Raj: https://github.com/uttoxi/Feynn-labs-project

Deeksha: https://github.com/Deeksha-45/mcdonals

Prashant Kumar: https://github.com/iitkgpian/Macdonalds_case

Rucha Bhide: https://github.com/wabi-sabi-wasabi/CaseStudyFeynn