Market Segmentation – A Case Study

by Mehul Bisht

Contents

Market Segmentation- A Case Study

I. Introduction

1. Market Segmentation

1.1   What is Market Segmentation?

Market Segmentation is the process of identifying segments of the target market and then dividing that target market into subgroups based on different factors such as demographics, needs, priorities, common interests, and other psychographic and/or behavioural criteria used to understand the target audience.

1.2 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. This matching process benefits consumers and suppliers and drives an organisation's marketing planning process.

A marketing plan consists of two components:
- A Strategic marketing plans
- A Tactical marketing plans

1.2.1 Strategic Marketing Plan:

A strategic marketing plan is a comprehensive document that outlines the goals, objectives, strategies, and tactics for a company's marketing activities. It provides a roadmap and direction for the marketing team, guiding their actions to achieve specific business outcomes. A plan of action or policy designed to achieve a major or overall aim. A long-range blueprint of an organization's expected image of destination. It is high-risk and proactive.

Example: Market Expansion

Objective: Enter new markets and increase market share

Strategies:
- Conduct market research to identify potential new markets and their characteristics.
- Adapt the product or service to meet the specific needs of the new markets.
- Develop localized marketing strategies that resonate with the target audience in each market
- Identify and collaborate with local distribution partners or retailers.
- Implement targeted advertising and promotional campaigns in the new market.
- Establish strategic alliances or partnerships with local businesses.
- Continuously track market penetration, customer acquisition, and revenue growth in the new markets.

1.2.2 Tactical Marketing Plan:

A tactical marketing plan refers to the specific actions, initiatives, and tactics employed to execute the strategies outlined in a broader marketing plan. It focuses on the operational details and day-to-day activities that drive marketing efforts towards achieving specific objectives and goals. In simple terms, a tactical marketing plan provides a detailed roadmap of the specific tasks, timelines and resources required to implement the marketing strategies defined in the overall marketing plan. A carefully planned action made to achieve a specific objective has low risks and is reactive.

Example: Content Creation and Marketing

- Developing and sharing relevant and valuable content that addresses the specific needs and interests of the target segment. This could include blog posts, articles, videos, or infographics.
- Leveraging email marketing to deliver personalized content and offers to segment specific audiences.
- Implementing search engine optimization (SEO) strategies to ensure that content is discoverable by the target segment when searching for relevant information.

The combination of good strategic marketing and good tactical marketing leads to the best possible outcome. Bad strategic marketing combined with bad tactical marketing leads to failure, but this failure unfolds slowly. A faster pathway to failure is to have excellent tactical marketing based on bad strategic marketing. This is equivalent to running full speed up to the wrong mountain. Good strategic marketing combined with bad tactical marketing ensures survival, albeit not in a particularly happy place.

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.
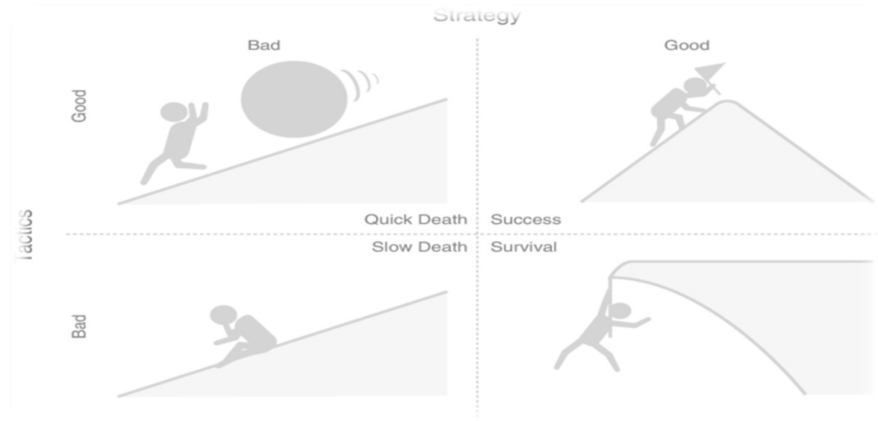
1.3 Benefits of Market Segmentation

- **Increased resource efficiency:** Marketing segmentation allows management to focus on certain demographics or customers. Instead of trying to promote the products to the entire market, marketing segmentation allows a focused, precise approach that often costs less compared to a broad-reach approach.

- **Strong brand image:** The marketing segment forces management to consider how it wants to be perceived by a specific group of people, once the market segment is identified, management must then consider what message to craft.

- **Greater potential for brand loyalty:** Marketing segmentation increases the opportunity for consumers to build long-term relationships with a company. A more direct, personal marketing approach may resonate with customers and foster a sense of inclusion, community, and a sense of belonging. In addition, market segmentation increases the probability that you land the right client that fits your product line and demographic.

- **Stronger market differentiation:** Market segmentation allows a company to pinpoint the exact message the way to convey to the market and competitors. This can also help in product differentiation by communicating specifically how a company is different from its competitors.

- **Better-targeted digital advertising:** Marketing segmentation enables a company to perform better-targeted advertising strategies. This includes marketing plans that direct efforts towards specific ages, locations, or habits via social media.

2. Market Segmentation Analysis

2.1 Market Segmentation Analysis Step-by-Step

Commonsense segmentation	Data-driven segmentation
STEP 1 - Deciding (not) to segment	
Is the market suitable? Can you make a long-term commitment?	Is the market suitable? Can you make a long-term commitment?
STEP 2 - Specifying the ideal target segment	
What would your ideal target segment look like?	What would your ideal target segment look like?
STEP 3 - Collecting data	
Collect data (segmentation variable and descriptor variables)	Collect data (segmentation and descriptor variables)
STEP 4 - Exploring data	
Explore data, pre-process if required.	Explore data, pre-process if required.
STEP 5 - Extracting segments	
Split consumers into segments using the segmentation variable.	Use distance-based, model-based or hybrid algorithms.
STEP 6 - Profiling segments	
--	Determine key features of the extracted market segments.
STEP 7 - Describing segments	
Describe segments in detail.	Describe segments in detail.
STEP 8 - Selecting (the) target segment(s)	
Evaluate segments and select target segment(s).	Evaluate segments and select target segment(s).
STEP 9 - Customising the marketing mix	
Develop a customised marketing mix.	Develop a customised marketing mix.
STEP 10 - Evaluation and monitoring	
Evaluate success, monitor changes.	Evaluate success, monitor changes.

II. Ten Steps of Market Segmentation Analysis

3. Step-1: Deciding (not) to Segment

3.1 Implications of Committing to Market Segmentation:

- Long-term commitment is necessary for market segmentation strategy
- Substantial changes and investments are required
- Costs of implementation, including research, surveys, and design
- Organizational adjustments to focus on market segments
- Decision-making at the executive level is crucial
- Systematic communication and reinforcement across all levels and units.

3.2 Implementation Barriers:

3.2.1 Barriers related to senior management:
- Lack of leadership, proactive championing, commitment, and involvement
- Insufficient allocation of resources for segmentation analysis and implementation

3.2.2 Barriers related to organizational culture:
- Lack of market or consumer orientation
- Resistance to change and new ideas
- Lack of creative thinking
- Poor communication and lack of information sharing across units
- Short-term thinking and unwillingness to make changes
- Office politics

3.2.3 Barriers related to training and expertise:
- Lack of understanding of market segmentation foundation and consequences
- Absence of a qualified marketing expert or formal marketing function
- Lack of a qualified data manager and analyst

3.2.4 Objective restrictions faced by the organization:
- Lack of financial resources
- Inability to make required structural changes

3.2.5 Process-related barriers:
- Unclear objectives of the segmentation exercise
- Lack of planning or inadequate planning
- Absence of structured processes and allocation of responsibilities

- Time pressure hindering the search for the best segmentation outcome

3.2.6 Operational level challenges:
- Resistance to using management techniques that are not understood
- Need for easy-to-understand market segmentation analysis and results presentation.

3.2.7 Considerations for overcoming barriers:
- Proactively identifying and removing barriers
- Serious considerations of abandoning segmentation if barriers persist
- Requirement for a resolute sense of purpose, dedication, patience, and problem-solving when implementing conclusions.

4. <u>Step 2: Specifying the Ideal Target Segment</u>

4.1 Segment Evaluation Criteria:

- **Size and Growth Potential**: Evaluate the size of each segment and its growth potential in terms of customer base, market value, or industry trends. Prioritize segments that are sizable and show potential for future growth
- **Profitability**: Assess the profitability of each segment by considering factors such as purchasing power, willingness to pay, and potential profit margins. Focus on segments that offer attractive revenue opportunities.
- **Customer needs and Preferences**: Analyse the unique needs, preferences, and pain points of each segment Prioritize segments that align closely with your business's products, capabilities, and value propositions.
- **Market Access**: Consider the accessibility and reachability of each segment. Evaluate whether you have the means to effectively target and serve the segment through existing distribution channels, marketing platforms, or geographic proximity.
- **Competitive landscape**: Assess the competitive intensity within each segment. Analyse the number and strength of competitors targeting the same segment and evaluate your ability to differentiate and gain a competitive advantage.
- **Compatibility with Business Objectives**: Determine how well each segment aligns with your business objectives, mission, and long-term strategic goals. Prioritize segments that contribute to your overall business direction and values.

4.2 Knock Out Criteria:

Segmentation work requires several criteria to be satisfied. If not, you should not commit:

- The segment must be homogeneous. Consumers in the segment must be similar to one another.
- The segment must be distinct. Consumers in the segment must be distinctly different in required characteristics or behaviour from people in other segments.
- The segment must be large enough. The segment must contain enough consumers to make it worthwhile to spend extra money on customizing the marketing mix.
- Consumers in the segment must match the strengths of the organizations. The organization must have the capability to satisfy the needs of consumers in the segment
- Consumers in the segment must be identifiable. It must be possible to spot them in the marketplace.
- The segment must be reachable. There must be a way to reach consumers in the segment to make the customized marketing mix accessible to them.

These criteria must be understood by senior management, the segmentation team, and the advisory committee. Most of the criteria do not require further specification, but some do. For example, while the size is non-negotiable, the exact minimum variable target segment size needs to be specified.

4.3 Criteria For Evaluating Segment Attractiveness:

The segmentation team must ask several questions which fall into two broad categories. Answering these two questions forms the based on the target decision

- Which of the market segments would the organization most like to target? Which segment would the organization like to commit to?
- Which of the organizations offering the same product would each of the segments most like to buy from? How likely is it that our organization would be chosen? How likely is it that each segment would commit to us?

Unfortunately, there is no standard formula for evaluating the attractiveness of market segments so a good deal of subjective judgement must be exercised. Nevertheless, several considerations can be used to assist in evaluating market segments for overall attractiveness.

4.4 Implementing a Structured Process

Specifying the Ideal target Segment means following a systematic and organized approach to identify the best target segment for your business. It involves:

- Gathering and reviewing market research data.

- Defining clear evaluation criteria based on your business goals.
- Analysing and ranking potential segments based on the criteria.
- Creating detailed profiles for the top-ranked segments.
- Validating findings through discussion and feedback.
- Selecting the ideal target segment.
- Developing a tailored positioning and value proposition.
- Implementing targeted marketing strategies and campaigns.
- Monitoring performance and adjusting as needed.

By following this structured process, you can make informed decisions and focus your marketing efforts on the most promising segment that aligns with your business objectives.

5. <u>Step 3: Collecting Data</u>

5.1 Segmentation Variables

- Empirical data is essential in both commonsense and data-driven market segmentation.
- In commonsense segmentation, a single characteristic (e.g., Gender) is used as the segmentation variable to split the sample into segments
- Other personal characteristics (e.g., age, number of vacations, benefits sought) serve as descriptor variables to describe the segments in detail.
- Data-driven market segmentation involves using multiple segmentation variables to identify or create market segments.
- The quality of empirical data is crucial for developing valid segmentation solutions and correctly assigning individuals to segments.
- Good data quality enables accurate segment descriptions, leading to effective product customization, pricing strategies, distribution channels, and advertising.
- Empirical data for segmentation studies can come from various sources such as surveys, observations (e.g., scanner data), or experiments.
- Survey data is common but may be unreliable for reflecting behaviour, especially socially desirable actions.
- It is important to explore multiple data sources and prioritize those that closely reflect actual consumer behaviour.

5.2 Segmentation Criteria

- The choice of segmentation criterion is a crucial decision that must be made by the organization before segment extraction and data collection.

- The term "segmentation criterion" refers to the nature of the information used for market segmentation, whereas "segmentation variable" refers to a specific measured value.
- Common segmentation criteria include geographic, socio-demographic, psychographic and behavioural factors.
- The decision of which segmentation criterion to use requires prior knowledge of the market and cannot be easily outsourced to consultants or data analysts.
- Relevant differences between consumers for segmentation purposes may include profitability, bargaining power, preferences for benefits or products, barriers to choose, and consumer interaction effects.
- While there are many segmentation criteria available, the recommendation is to use the simplest approach that is effective for the specific marketing context.
- The principle of "doing the least you can" is often advised, meaning if that demographic or geographic segmentation is sufficient for the product or service, there is no need to opt for complex criteria like psychographic segmentation.
- The effectiveness of a segmentation criterion should be evaluated based on its suitability for the product or service and its cost-effectiveness.
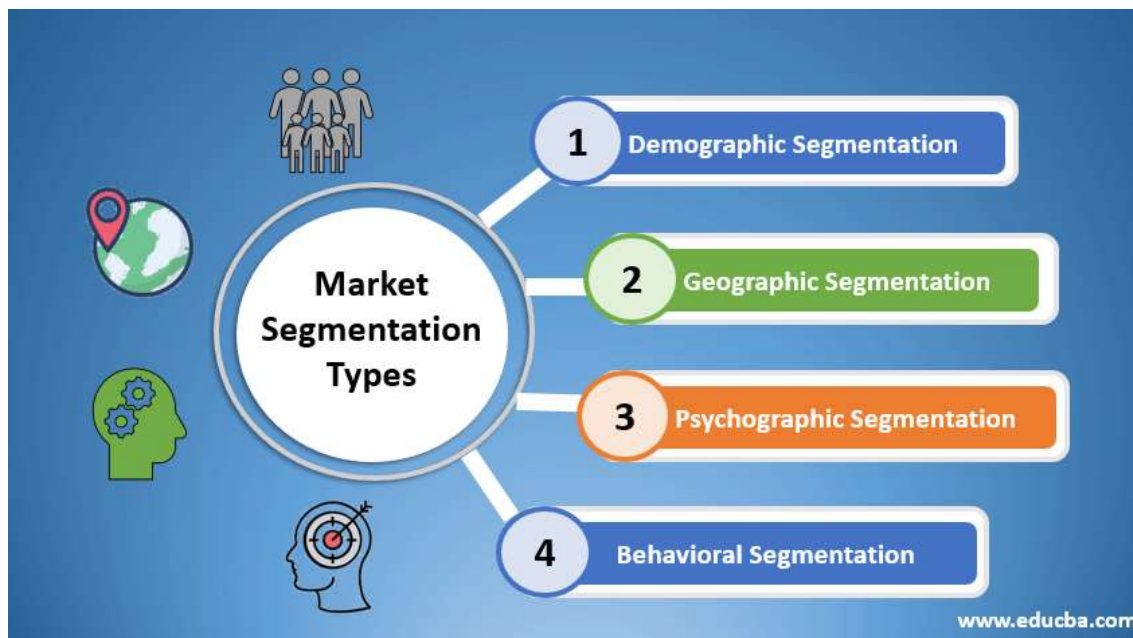
5.2.1 Geographic Segmentation:

Geographic segmentation, which uses the consumer's location of residence as the segmentation criterion, is often the original and the simplest approach to market segmentation. Here are some key points about geographic segmentation:

Advantages:

14

- Easy assignment of consumers to geographic units, allowing for easy targeting of communication messages and selection of appropriate communication channels.
- Pragmatic reasons, such as language differences or country-specific offerings, can make geographic segmentation a relevant approach, especially for global companies.

Disadvantages:
- Living in the same geographic area does not necessarily mean that people share other relevant characteristics for marketers, such as product preferences or benefits sought.
- Socio-demographic criteria often play a more significant role in explaining differences in consumer behaviour within a geographic area.
- People from the same country or area can have diverse preferences and needs, making geographic segmentation alone insufficient for effective targeting.

5.2.2 Socio-Demographic Segmentation

Socio-demographic segmentation criteria such as age, gender, income, and education are commonly used in market segmentation. Some key points about socio-demographic segmentation are:

Advantages:
- Easily identification of segment membership for each consumer based on readily available socio-demographic information.
- Socio-demographic criteria can be relevant and useful in certain industries where consumer preferences are associated with specific socio-demographic characteristics.
- Socio-demographic factors can provide some explanatory power for specific product preferences or behaviours, such as the choice of luxury goods, cosmetics, baby products, retirement villages or tourism resort products.

Disadvantages:
- Socio-demographic criteria alone may not fully explain or capture the underlying reasons for product preferences or consumer behaviour. Other factors like values, tastes, and preferences often have a stronger influence on buying decisions.
- Socio-demographic segmentation may overlook important differences between segments that have similar socio-demographic characteristics.
- Demographics alone typically account for a relatively small percentage (e.g., around 5%) of the variance in consumer behaviour, according to an estimate by Haley (1985)

5.2.3 Psychographic Segmentation:

Psychographic segmentation involves grouping individuals based on psychological criteria, such as beliefs, interests, preferences, aspirations, or benefits sought when purchasing a product.
Some key points about psychographic segmentation are:

- **Scope of psychographics:** Psychographics encompass measures of the mind and various psychological dimensions that influence consumer behaviour. Benefits segmentation and lifestyle segmentation are two popular approaches to psychographic segmentation.
- **Complexity of Psychographic Criteria:** Psychographic criteria are more complex than geographic or socio-demographic criteria because they aim to capture individuals' psychological characteristics. It is challenging to find a single characteristic that adequately represents the desired psychographic dimension. As a result, multiple segmentation variables, such as different travel motives or perceived risks, are often used in psychographic segmentation studies.
- **Reflecting Consumer Behaviour:** The psychographic approach offers the advantage of being more reflective of the underlying reasons for differences in consumer behaviour. By understanding consumers' beliefs, interests, and preferences, marketers, can tailor their offerings to better meet their needs and motivations. For example, using travel motives as a basis for psychographic segmentation in the tourism industry can help identify distinct segments with different preferences for cultural, adventure, or relaxation-oriented vacations.
- **Challenges and consideration:** Psychographic segmentation poses challenges in determining segment membership for consumers due to the complexity of psychological factors involved. Additionally, the effectiveness of psychographic segmentation relies on the reliability and validity of the measures used to capture the desired psychographic dimensions. Accurate and role robust data collection is crucial for obtaining meaningful psychographic insights.

5.2.4 Behavioural Segmentation

The approach of segmentation based on similarities in behaviour or reported behaviour is known as behavioural segmentation. This approach relies on various behavioural indicators to group individuals into segments. Here are some key points about behavioural segmentation:

- **Behavioural indicators:** Behavioural segmentation utilizes a wide range of behaviours or reported behaviours as segmentation variables. These can

include prior product experience, purchase frequency, purchase amount, information search behaviour and brand choice behaviour over time. By analysing these behaviours, marketers can identify patterns and similarities among consumers.

- **Superiority of Behavioural Criteria:** In some cases, behavioural criteria are superior to other segmentation variables. For example, in the context of tourism, behaviours reported by tourists were found to be more effective than geographic variables in segmenting the market, This suggests that analysing actual behaviour can provide valuable insights for segmentation.

- **Basis in Actual Behaviour:** The advantage of behavioural segmentation is that it directly uses the behaviour of interest as a basis for segment extraction. By focusing on actual behaviour rather than stated or intended behaviour, marketers can create segments based on the most relevant similarities. This approach can lead to more accurate targeting and personalized strategies.

- **Availability of Behavioural Data**: One challenge with behavioural segmentation is the availability of relevant behavioural data. It may be difficult to access comprehensive behavioural data, especially when trying to include potential customers who have not previously purchased the product. This limitation can restrict the scope of behavioural segmentation to existing customers rather than the entire target market.

- **Avoiding the Need for Psychological Measures:** Behavioural segmentation offers an advantage by not requiring the development of measures for psychological constructs. Instead of relying on psychological variables, marketers can directly analyse observable behaviours to identify meaningful segments.

5.3 Data from Survey Studies

5.3.1 Choice of variable:

The selection of variables in market segmentation analysis is crucial to ensure the quality of the segmentation solution. Here are some important points to note:

- **Data-Driven Segmentation:** In data-driven segmentation, it is important to include all variables that are relevant to the construct captured by the segmentation criterion. This ensures that the analysis captures the necessary information for accurate segmentation. However, unnecessary variables should be avoided to prevent respondent fatigue and to maintain the quality of responses.

- **Noisy Variables:** Noisy variables, also known as masking variables, do not contribute any meaningful information to the identification of the market segments. Their presence can hinder the algorithm's ability to extract the correct segmentation solution. Noisy variables can result from poorly developed survey questions or the inclusion of unnecessary variables. Careful attention should be given to selecting variables that are informative and contribute to the segmentation analysis.

- **Avoiding Redundancy:** Redundant questions or items in a survey can be problematic in market segmentation analysis. They interfere with the algorithms' ability to identify accurate segmentation solutions. It is recommended to avoid including redundant questions and to focus on asking necessary and unique questions. Redundancy often arises from traditional psychometric scale development principles, but recent research suggests questioning this practice in the context of market segmentation analysis.

- **Exploratory Research:** Developing a good questionnaire for market segmentation typically involves conducting exploratory or qualitative research. Exploratory research provides insights into people's beliefs and perspectives that may not be captured through survey research alone. These insights can be categorized and included as answer options in the questionnaire, ensuring that important variables are not omitted.

5.3.2 Response Option:

Different types of response options have implications for the type of data and distance measures that can be used in segmentation analysis. Here are the key considerations for each type of response option:

- **Binary/Dichotomous Data:** When respondents can choose only one of two options, such as "yes" or "no", the data generated is binary. Binary data can be represented as 0s and 1s, and the distance between the two values is clearly defined. Binary data poses no difficulties for subsequent segmentation analysis.

- **Nominal Data:** When respondents select an answer from a range of unordered categories, such as selecting their occupation from a list, the data generated o is nominal. Nominal variables can be transformed into binary data by introducing a binary variable for each answer option. This transformation allows for further analysis using binary data.

- **Metric Data:** When respondents provide numerical information, such as the age or the number of nights stayed at the hotel, the data generated is metric. Metric data can be used in any statistical procedure, including distance-based measures, making them well-suited for segmentation analysis.

- **Ordinal Data:** The most common response format in surveys involves respondents rating their agreement with statements using a limited no. of order options (e.g., a Likert Sale). This generates ordinal data, where the options are ordered, but the distance between adjacent options is not clearly defined. Applying standard distance measures to ordinal data requires making strong assumptions. Properly addressing the distance measure issue is discussed in Step 5 of the segmentation analysis.

5.3.3 Response Style:

Response biases and styles can significantly impact the results of segmentation analysis. Here are some key points to consider regarding response biases and their implications for segmentation:

- **Response Biases:** Response biases refer to systematic tendencies of respondents to answer questionnaire items based on factors other than the specific content of the items. Examples of response biases include acquiescence bias (tendency to agree with all questions), extreme response style (tendency to use extreme answer options), and midpoint preference (tendency to choose the midpoint option).

- **Impact on Segmentation:** Response biases can affect segmentation results because they distort the true underlying beliefs or preferences of respondents. Segmentation algorithms cannot distinguish between genuine beliefs and response biases, potentially leading to misinterpretation of market segments. For example, a segment that appears to be high-spending based on positive responses to all spending-related questions may reflect an acquiescence bias rather than actual high-spending behaviour.

- **Minimizing Response Styles:** Minimizing the risk of capturing response styles is crucial for accurate segmentation. Steps that can be taken to mitigate response biases include:
  a) Designing a balanced questionnaire with both positively and negatively worded items.
  b) Using multiple response formats (e.g., a Likert Scale, ranking scales) to reduce acquiescence bias.
  c) Employing response validation techniques, such as including attention-check questions or using psychometric scales to identify response styles.

- **Additional Analysis and Exclusion:** if attractive market segments emerge that potentially exhibit response patterns influenced by response styles, additional analyses are necessary. These may involve identifying and excluding respondents affected by response biases from the target segment. Alternatively, conducting further analysis to differentiate

between genuine segment characteristics and response styles can help ensure accurate segmentation results.

6. <u>Step 4: Exploring Data</u>

6.1 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, income, education, occupation, and various travel motives

- The summary of the data set provides insights into the characteristics of variables, such as the number of respondents by gender, the age distribution (minimum, maximum, quartiles) and the presence of missing data (Nas).

6.2 Data Cleaning

- Data cleaning is the first step before data analysis, involving checking for correct values and consistent labels for categorical variables.

- A plausible range 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 data analysis, as documenting and saving the steps taken allows for future replication and enables continuous monitoring and analysis of segmentation solutions.

6.3 Descriptive Analysis:

- **Descriptive Analysis:** It involves summarizing and understanding the data by using numerical and graphical representation. 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 the 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. The histogram displays the frequency of observation within specific value ranges (bins).

- **Boxplots in R:** Boxplots summarize the distribution of numeric variables. R's boxplot () functions generate a box-and-whisker plot, which includes the minimum, first quartile, median, third quartile, and maximum value. The plot helps identify skewness, outliers, and quartile range. Whiskers are typically limited in length to avoid emphasizing outliers.

6.4 Preprocessing

6.4.1 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 analyse.

- Categorical variables can be converted to numeric variables if the distance between adjacent scale points on the ordinal scale is 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 procedure works 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 a comparable impact on segmentation results.

- Alternative standardization methods may be necessary when data contains outliers. Robust estimates, such as the median and the interquartile range, can be used instead of the mean and standard deviation to minimize the influence of outliers.

6.4.2 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 for 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 deviation of the principal components, which reflects their importance and the rotation matrix, which specifies how the original variables contribute to each principal component.

- PCA can be useful for exploring data, identifying highly correlated variables, and potentially removing redundant variables from the segmentation base.

7. <u>Step-5: Extracting Segments</u>

Market segmentation analysis using data-driven approaches is exploratory, as consumer datasets are typically unstructured and diverse. 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. Different algorithms can impose different structures on the segments, no single algorithm is universally superior, and the selection of extraction methods should consider the data characteristics and desired segment characteristics.

This chapter provides an overview of popular extraction methods, including distance-based and model-based approaches, as well as methods that combine multiple aims. Comparing alternative segmentation solutions and considering data characteristics are crucial for obtaining a robust final solution. Factors such as sample size, scale levels of variables, and expected segment characteristics guide algorithm selection. Additionally, the treatment of directly observable and indirectly accessible characteristics, as well as binary segmentation variables, must be considered.

7.1 Distance-Based Methods

Market segmentation aims at grouping consumers into groups with similar needs or behaviours. Different distance measures can be used to determine the similarity and dissimilarity between the consumers.

Euclidean distance calculates the straight-line distance between two points in multidimensional space. It considers all dimensions of the vectors and calculates the square root of the sum of squared differences between corresponding elements.

Manhattan distance calculates the distance between two points by summing the absolute differences between corresponding elements. It also considers all dimensions of the vectors.

Asymmetric binary distance measure is specific to binary vectors, where elements are either 0 or 1. It calculates the proportion of common 1s overall dimensions where at least one vector contains a 1.

7.1.1 Hierarchical Methods

Hierarchical clustering is an exploratory technique that offers a structured approach to grouping data. It provides a sequence of nested partitions. There are two types of hierarchical methods divisive and agglomerative.

In divisive hierarchical clustering, the process starts with the complete dataset and splits it into two segments, each segment is further split into two, and this process continues until each observation has its segment. On the other hand, agglomerative hierarchical clustering takes the opposite approach. It starts with each observation representing its segment (singleton clusters) and gradually merges the closest segments until the entire dataset forms one large segment.

This can be performed using different distance measures and linkage methods. Dendrograms are commonly used to visualize the clustering result. Dendrograms visualize the hierarchy of segments formed at each step of the clustering process, with the height of the branches indicating the distance between the clusters. Dendrograms can be used as a guide to select the number of market segments, although they may not always provide clear guidance in consumer data analysis due to the complexity of the data.

7.1.2 Partitioning Methods

Hierarchical clustering methods are often used for the analysis of small data sets up to a few hundred observations. They create a hierarchical structure of clusters, represented by a dendrogram. However, for larger data sets, dendrograms become difficult to read and the pairwise distance matrix may not fit into computer memory. For larger data sets with more than 1000 observations, the partitioning method is k-means clustering, which divides the data into subsets or segments based on similarity.

7.1.2.1 k-Means and k-Centroid Clustering

k-means clustering is the most used distance-based partitioning clustering algorithm. Using random consumers from the datasets as starting points, the standard k-means clustering algorithm iteratively assigns all consumers to the cluster centres (centroids, segment representatives), and adjusts the location of the cluster centres until cluster centres do not change anymore. Standard k-means clustering uses the squared Euclidean distance. Generalisation Using other distances is also referred to as k-centroid clustering.

The k-means algorithm involves several steps:

- Specify the desired number of segments, denoted as k.
- Randomly select k observation as initials cluster centroids.
- Assign each observation to the closest centroid to form an initial partition.

- Recompute the cluster centroids based on other current partitions.
- Repeats steps 3 and 4 until convergence or a maximum number of iterations is reached.

7.1.2.2 Improved k-means:

In market segmentation analysis, various algorithms and methods can be used to refine and improve the k-means clustering algorithm. One common improvement is to initialize the k-means algorithm with the smart starting values instead of randomly selecting consumers from the data set. The best starting point is those that effectively, represent the data, with representatives that are close to their segment members. This approach helps avoid the problem of getting trapped in local optima.

7.1.2.3 Hard Competitive Learning:

This method differs from k-means in how segments are extracted. Instead of using all consumers in the data set at each iteration, hard competitive learning randomly selects one consumer and moves its closest segment representative towards it. The procedural difference can lead to different segmentation solutions compared to k-means, and it may find the globally optimal solution while k-means get stuck in a local optimum.

7.1.2.4 Neural Gas and Topology Representing Networks:

Neural gas and topology representing networks are a further extension of hard competitive learning. Neural gas adjusts not only the closest representative but also the second closest representative toward the randomly selected consumer, with a smaller adjustment for the second closest. Topology representing networks counts how often each pair of segment representatives is closest and second closest to a randomly drawn consumer, creating a virtual map that represents the relationships between representatives.

7.1.2.5 Self-organizing maps:

Self-organizing maps are another variation of hard competitive learning that positions segment representatives on a regular grid. It uses a rectangular or hexagonal grid and adjusts the representatives based on the selected random consumer and its closest neighbours. The advantage of self-organizing maps is that the numbering of market segments aligns with the grid, providing a structured output. However, the sum of distances between segment members and representatives can be larger compared to other clustering algorithms due to the restrictions imposed by the grid.

7.1.2.6 Neural Networks:

Neural networks, specifically auto-encoding neural networks, are a different approach to cluster analysis. They use a single hidden layer perceptron to learn the best representation of the data and predict the inputs as accurately as possible. The parameters connecting the hidden layer to the output layer can be interpreted as a segment representative, while the parameters connecting the input layer to the hidden layer indicate membership in different segments.

7.1.3 Hybrid Approaches:

Hybrid segmentation approaches aim to leverage the strength of both hierarchical and partitioning algorithms. They begin with a partitioning algorithm for scalability, then transition to hierarchical clustering using reduced data to determine the appropriate number of segments. This combination allows for efficient segmentation of large datasets while enabling visualization and decision-making based on segment similarities.

7.1.3.1 Two-Step Clustering:

This process consists of two steps: a partitioning procedure followed by a hierarchical procedure. In the first step, a partitioning algorithm (such as k-means) is applied to the data to reduce the size and extract representative members from each cluster. The number of clusters extracted in this step is not crucial and can be larger than the actual number of segments sought. In the second step, a hierarchical cluster analysis is performed using the cluster segments and centres and segment sizes obtained from the first step. The resulting dendrogram helps identify the number of market segments. Finally, the original data is linked to the segmentation solution derived from the hierarchical analysis.

7.1.3.2 Bagged Clustering:

Bagged clustering, on the other hand, combines hierarchical clustering and partitioning clustering with bootstrapping. It starts by creating multiple bootstrap samples from original datasets. Each bootstrap sample is then clustered using a partitioning algorithm. The cluster centroids obtained from these repeated partitioning analyses serve as the data set for the hierarchical clustering step. The final segmentation solution is determined by selecting a cut point in the dendrogram and assigning each observation to the closest market segment.

7.2   Model-Based Methods:

Model-based methods offer an alternative approach to market segmentation analysis by using a finite mixture model. The models capture segment-specific characteristics and sizes, and various statistical techniques are employed to estimate the model parameters and assign consumers to segments.

7.2.1 Finite Mixture Distributions:

The finite mixture model is represented by a combination of segment-specific models, where each segment is associated with a set of parameters. The parameters, including segment sizes and segment-specific characteristics, need to be estimated using statistical estimation techniques such as maximum likelihood estimation or Bayesian inference. To assign consumers to segments, probabilities of segment memberships are based on consumer information and the estimated parameter values. The segment with the highest probability is then assigned to the consumer.

7.2.1.1 Normal Distributions:

A mixture of the normal distribution is suitable for, market segmentation when the segmentation variables are metric, such as money spent on different consumption categories time spent engaging in different vacation activities, or body measurements for different clothing sizes. The multivariate normal distribution is suitable for modelling covariance between variables, and it commonly occurs in biological and business contexts. For instance, physical measurements of humans, such as height, arm length, leg length, and foot length, can be well approximated by a multivariate normal distribution.

If there are p segmentation variables used, then there will be p mean values and each segment will have a segment-specific mean vector of length p. In addition to the variance of the p segmentation variable, the covariance structure can also be modelled. This results in a p x p covariance matrix for each segment. The covariance matrix contains the variances of the p segmentation variables on the diagonal and the covariance between pairs of segmentation variables in other entries. The covariance matrix is symmetric and contains $p(p+1)/2$ unique values.

The segment-specific parameters are the combination of the mean vector and the covariance matrix. The number of parameters to estimate is $p+(p+1)/2$ accounting for the mean vector and the unique values in the covariance matrix.

7.2.1.2 Binary Distributions:

The finite mixtures of binary distributions model market segmentation based on binary segmentation variables representing customers' activities. In this approach, binary segmentation variables are used to represent customer preferences or activities, where a value of 1 indicates engagement

in a specific activity and 0 indicates non-engagement. The parameters of the segmentation-specific models, which represent the probabilities of observing a 1 in each variable, are extracted. These probabilities characterize the segments and can be used to create segment profiles. Overall, a mixture of binary distributions provides a way to model the association between binary variables and identify distinct market segments based on activity patterns.

7.2.2 Finite Mixtures of Regressions:

Finite mixtures of regressions provide a different perspective on market segmentation compared to distance-based clustering methods. They analyse the relationship between variables and allow for the identification of distinct segments with varying regression patterns. A finite mixture of regression models assumes the existence of a dependent target variable y that can be explained by a set of independent variables x. The functional relationship between the dependent and independent variables is considered for different market segments.

7.3 Algorithms with Integrated Variable Selection:

The section highlights the importance of variable selection in segmentation analysis. While many segmentation algorithms assume that all variables contribute to determining the segmentation solution, this may not always be the case. In situations where the segmentation variables contain redundant or noisy information, it becomes necessary to identify and select suitable variables for the analysis. Variable selections play a crucial role in segmentation analysis, especially when dealing with redundant or noisy variables. Different algorithms, such as bi-clustering, VSBD, and factor-cluster analysis, offer integrated approaches to segment extraction while simultaneously selecting suitable segmentation variables, considering the specific characteristics of binary data.

7.3.1 Biclustering Algorithms:

Biclustering is a method for simultaneously clustering consumers and variables in market segmentation analysis. The Biclusteing algorithm can be applied to different types of data, including binary data. In the binary case, Biclustering aims to extract market segments consisting of consumers who have a value of 1 for a specific group of variables. Biclustering offers advantages in market segmentation with many variables. It avoids data transformation, which can introduce bias and allows for the identification of niche markets by setting specific control parameters. However,

Biclustering methods may not group all consumers, leaving some ungrouped individuals who do not fit into any segment.

Biclustering is a powerful approach for market segment analysis, particularly with binary data and many variables. It enables the identification of groups of consumers and variables with common patterns, providing insights into niche markets and avoiding data transformation biases.

7.3.2 Variable Selection Procedure for Clustering Binary Data (VSBD):

The VSBD method by Brusco is a variable selection procedure for clustering binary data. It uses the k-means algorithm and within-cluster sum-of-squares criterion to identify relevant variables and remove masking variables. The procedure involves an iterative process of adding variables based on their impact on the clustering solution, and it requires specifying the number of segments in advance.

7.3.3 Variable Reduction: Factor Cluster Analysis:

The factor-cluster analysis is a two-step procedure used for data-driven market segmentation analysis. In the first step, the segmentation variables are subjected to factor analysis, and the raw data is discarded. In the second step, market segments are extracted using the factor scores obtained from the factor analysis. Factor-cluster analysis is often used when the number of segmentation variables is too high relative to the sample size. Factor-cluster analysis lacks conceptual justification and can result in a loss of information, data transformation, and difficulties in interpretation. It is generally recommended to perform cluster analysis on raw data rather than relying on factor scores for market segmentation purposes.

7.4 Data Structure Analysis:

Data structure analysis in market segmentation is aimed at assessing the reliability and stability of segmentation solutions, rather than determining an optimal solution with a clear criterion. Since it is not feasible to validate multiple segmentation strategies simultaneously, validation in market segmentation typically focuses on evaluating the stability of solutions across repeated calculations.

The purpose of data structure analysis is to gain insights into the properties of the data and guide methodological decisions. It helps determine whether natural, distinct, and well-separated market segments exist in the data. If such segments exist, they can be easily identified. If not, analysts need to explore

various alternative solutions to identify the most useful segment(s) for the organization.

There are four main approaches to data structure analysis:

7.4.1 Cluster indices:

Cluster indices provide a measure of within-cluster homogeneity and between-cluster separation. These indices help assess the quality of segmentation solutions and identify the number of segments that best fit the data.

7.4.2 Gorge Plots:

Gorge plots visually represent the stability of solutions by plotting the average within-cluster dissimilarity as the number of segments increases. Gorge plots can reveal the presence of well-separated segments and help determine the appropriate number of segments.

7.4.3 Global stability analysis:

Global stability analysis examines the overall stability of segmentation solutions across multiple runs with slightly modified data or algorithms. It provides insights into the robustness of the identified segments and helps assess the reliability of the results.

7.4.4 Segment level stability analysis:

Segment-level stability analysis focuses on the stability of individual segments across different runs. It examines the consistency of segment membership and characteristics, allowing for a more detailed understanding of the stability and reliability of the segmentation solution.

These approaches collectively contribute to data structure analysis and assist in making informed decisions about the number of segments to extract and the reliability of the segmentation results. By assessing the stability and structure of the data, analysts can gain valuable insights and choose the most appropriate segmentation solution for their organization.

8 Step 6: Profiling Segment

8.1 Identifying Key Characteristics Segments

- Profiling is essential in data-driven market segmentation to understand the characteristics and defining features of resulting market segments.

- Profiling is not required in common sense segmentation, as the segments are predefined based on obvious characteristics like age group.
- profiling in data-driven segmentation aims to identify the defining characteristics of market segments based on the segmentation variables.
- Profiling involves characterizing the market segments individually and comparing them to other segments to understand their uniqueness.
- Data-driven market segmentation solutions can be challenging to interpret, and many managers struggle to understand the results, per the study by Dolnicar and Lazarevski (2009). A clear and concise presentation of segmentation results is crucial for effective decision-making.

8.2 Traditional Approaches to Profiling Market Segments

- Data-driven segmentation solutions are often presented in ways that can be challenging to interpret, such as high-level summaries that oversimplify segment characteristics or large tables with exact percentages for each segmentation variable.
- It illustrates the difficulty in interpreting segment characteristics based on exact percentages for each variable, requiring numerous comparisons between segments and the overall values. Profiling all segments in this manner can be a tedious and time-consuming task.
- Comparing multiple segmentation solutions further increases the complexity, as each solution may contain different segment definitions. In each case, the number of comparisons needed to understand the defining characteristics of the segments becomes significantly higher, making it a challenging task for users.

8.3 Segment Profiling with Visualizations

- Graphics and data visualizations play a crucial role in exploratory statistical analysis, including cluster analysis. They provide insights into the complex relationships between variables and make the interpretation of market segmentation results easier.
- Visualizations of segmentation solutions are valuable for inspecting segments in detail, interpreting segment profiles, and assessing the usefulness of different market segmentation solutions. They assist data analysts and users in making critical decisions when selecting the most appropriate segmentation solution.

8.4 Identifying Defining Characteristics of Market Segment

- Segment profile plots provide a visual representation of how each market segment differs from the overall sample across segmentation variables. They are a direct translation of tables and allow for a quick understanding of the defining characteristics of each segment.
- Visualization, such as segment profile plots, are easier and faster to interpret than tables, even when well-structured. They provide a comprehensive

overview of segment differences and interpret segmentation results more accessible.

- The order of segmentation variables in visualizations can be rearranged to improve clarity and facilitate interpretation. Variables can be ordered based on similarity of answer patterns, such as through hierarchical clustering of the variables.

- Marker variables in segment profile plots are highlighted in colour to indicate their significance in characterizing a segment. These variables have substantial differences in means compared to the overall sample, usually defined as deviating by more than 0.25 or 50% from the total mean.

- Visualizations, like segment profile plots, help in assessing the usefulness of a market segmentation solution and support the decision-making process of selecting the most appropriate solution. They enable data analysts and users to compare and evaluate different segment profiles.

- Eye tracking studies have shown that visualizations, such as segment profile plots, require less cognitive effort and processing time compared to tables. They allow for faster extraction of information, leading to easier interpretations and comprehension of segmentation results. Well-designed graphs offer a valuable return on investment, especially for managers making strategic decisions based on segmentation outcomes.

8.5 Assessing Segment Separation:

- Segment separation plots visualize the overlap of segments in the data space, providing an overview of the separation between segments, They are particularly useful when the number of segmentation variables is low but can become complex as the number of variables increases.

- In segment separation plots, the scatter plot displays the original data points coloured by segment memberships, while cluster hulls indicate the shape and spread of the true segments. Neighbourhood graphs show similarity between segments, with thicker lines indicating more observations sharing segment centres.

- For higher-dimensional data sets, projection techniques like principal component analysis can be used to reduce the dimensions and create segment separation plots. This allows for a visual representation of segment separation in a reduced-dimensional space.

- Segment separation plots can become cluttered and hard to Interpret due to overlapping segments, especially when there is limited separation between segments. Adjustments can be made, such as modifying colours, omitting observations, and highlighting the inner areas of each segment, to improve readability and interpretation.

- Segment separation plots visualize a specific projection of the data, and different projections may result in different separation patterns. It is important to consider multiple projections and not draw conclusions solely based on a single visualization. Interpretation should be based on the specific projection shown and the characteristics of the data.

9 Conclusion

It was noted in this study that market segmentation is regarded essential by marketing practitioners for various reasons, including targeting, proposition development, price formulation and development of mass communication. Though being conceptualised as simple in its rationale, the process of segmentation is not necessarily easy and it requires various considerations should be considered. From the literature, it is evident that many marketers are expressing concern about implementation and the integration of segmentation into marketing strategy. To address this, priorities in future segmentation research include the selection and incorporation of new variables into segmentation models, as well as the development of new and innovative segmentation strategies. Using market segmentation, companies can identify their target audiences and personalize marketing campaigns more effectively. This is why market segmentation is key to staying competitive. It allows you to understand your customers, anticipate their needs, and seize growth opportunities. This powerful technique allows you to improve your decision-making, and marketing efforts, and improve your company's bottom line. The key to successful market segmentation remains data quality; therefore, you need to pick your data provider after doing your due diligence, ensuring that you have access to the latest industry information in accessible and easy-to-understand formats.

10 References

[1] S. Dolnicar, B. Grun and F. Leisch, "Market Segmentation Analysis: Understanding it, doing it, and Making it Useful," Springer Singapore,2018

[2] https://rb.gy/lazcm9

[3] https://rb.gy/ehf02b

[4] https://rb.gy/gn7kuw