21 June 2023

Feynn Labs

Case Study Report

MARKET SEGMENTATION ANALYSIS

Manishankar Bag(Team Leader), Joy Saha, Manish Kumar, Soumyajit Mishra, Sudip Sahoo.

Contents

GitHub Link of Manishankar Bag : GitHub Link

GitHub Link of Joy Saha : GitHub Link

GitHub Link of Manish Kumar : GitHub Link

GitHub Link of Soumyajit Mishra : GitHub Link

GitHub Link of Sudip Sahoo : GitHub Link

Introduction

This portion is contributed by - Manishankar Bag

A) Strategic And Tactical Marketing:

Market segmentation is the process of dividing a larger market into smaller, distinct groups of consumers who share similar characteristics, needs, or behaviours. The goal of market segmentation is to identify specific segments within a target market that can be targeted with tailored marketing strategies and messages.

A marketing plan consists of two components:

- A strategic marketing plans.
- A tactical marketing plans.

B) 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 over all aim. A long range blueprint of an organization's expected image of destination. It has high risks 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 markets.
- Establish strategic alliances or partnerships with local businesses
- Continuously track market penetration, customer acquisition, and revenue growth in the new markets.

C) Tactic 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. It has low risks and 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.

D) Benefits Of Market Segmentation:

- **Increased resource efficiency:** Marketing segmentation allows management to focus on certain demographics or customers. Instead of trying to promote products to the entire market, marketing segmentation allows a focused, precise approach that often costs less compared to a broad reach approach.
- **Stronger brand image:** 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. Because this message is directed at a target audience, a company's branding and messaging is more likely to be very intentional. This may also have an indirect effect of causing better customer experiences with the company.
- **Greater potential for brand loyalty:** Marketing segmentation increases the opportunity for consumers to build long-term relationships with a company. More direct, personal marketing approaches 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 gives a company the opportunity to pinpoint the exact message that way to convey to the market and to competitors. This can also help in product differentiation by communicating specifically how a company is different from its competitors. Instead of a broad approach to marketing, management crafts a specific image that is more likely to be memorable and specific.
- **Better targeted digital advertising.** Marketing segmentation enables a company to perform better targeted advertising strategies. This includes marketing plans that direct effort towards specific ages, locations, or habits via social media.

Step 1: Deciding (not) to Segment

1. Implications of Committing to Market Segmentation

"Implications of Committing to Market Segmentation" explores the significance and consequences of embracing market segmentation as a strategic approach. By committing to market segmentation, businesses can effectively target specific customer groups, enhance marketing effectiveness, and maximize resource allocation. The chapter delves into the benefits and challenges associated with market segmentation, emphasizing the importance of aligning organizational capabilities, resources, and strategies with the identified market segments. It discusses how segmentation enables businesses to tailor their products, messages, and marketing efforts to better meet customer needs and preferences. Additionally, the chapter emphasizes the need for ongoing evaluation and adaptation to ensure the continued relevance and success of market segmentation strategies.

2 Implementation Barriers

Implementation Barriers highlights the challenges and obstacles that organizations may encounter when attempting to implement market segmentation strategies effectively. The chapter explores various barriers that can hinder successful implementation, such as resistance to change, lack of top management support, insufficient resources, and inadequate understanding of customer needs. It emphasizes the importance of overcoming these barriers through proactive planning, clear communication, and alignment of organizational objectives. The chapter also discusses potential solutions and best practices for addressing implementation barriers, including conducting thorough market research, fostering a culture of innovation, and investing in training and development. By addressing these barriers, businesses can enhance their ability to leverage market segmentation effectively and gain a competitive advantage.

Step 2: Specifying the Ideal Target Segment

This portion is contributed by - Sudip Sahoo

1.Segment Evaluation Criteria

Segment evaluation criteria are a set of factors used to assess the attractiveness and suitability of different market segments for an organization's marketing efforts. These criteria help in identifying segments with the highest potential for success and resource allocation. Commonly used segment evaluation criteria include size and growth potential, profitability, accessibility, compatibility, differentiation, stability, competitive intensity, feasibility, strategic fit, and return on investment.

Size and growth potential consider the segment's market size and future growth prospects. Profitability examines the segment's profit potential based on factors like customer purchasing power and cost of serving the segment. Accessibility assesses the ease of reaching and serving the segment. Compatibility evaluates the fit between the organization's offerings and the segment's needs. Differentiation considers the uniqueness of the segment's needs for targeted marketing. Stability assesses the segment's sustainability over time. Competitive intensity evaluates the level of competition within the segment. Feasibility considers the organization's capability to effectively serve the segment. Strategic fit examines the alignment between the segment and the organization's goals. Return on investment evaluates the potential returns compared to resources required.

By applying these evaluation criteria, organizations can prioritize segments that offer the greatest opportunities for success and make informed decisions regarding resource allocation and marketing strategies.

2.Knock-Out Criteria

Knock-out criteria, also known as exclusion criteria, are a set of factors or conditions used to eliminate or exclude certain market segments from further consideration in the segmentation process. These criteria help organizations narrow down their focus to the most relevant and viable segments. The purpose of knock-out criteria is to quickly identify segments that do not meet specific requirements or are not aligned with the organization's strategic objectives.

• The segment must be homogeneous; members of the segment must be similar to one another.

• The segment must be distinct; members of the segment must be distinctly different from members of other segments.

• The segment must be large enough; the segment must contain enough consumers to make it worthwhile to spend extra money on customising the marketing mix for them.

• The segment must be matching the strengths of the organisation; the organisation must have the capability to satisfy segment members' needs.

• Members of the segment must be identifiable; it must be possible to spot them in the marketplace. • The segment must be reachable; there has to be a way to get in touch with members of the segment in order to make the customised marketing mix accessible to them.

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

3. Attractiveness Criteria

Attractiveness criteria are factors used to assess the desirability and potential of market segments. These criteria include size and growth potential, profitability, accessibility, differentiation, stability, and strategic fit. By evaluating segments based on these criteria, organizations can prioritize and target the most promising segments for their marketing efforts.

4. Implementing a Structured Process

Implementing a structured process is crucial when conducting market segmentation analysis. This process involves following a systematic and organized approach to ensure effectiveness and efficiency in segment identification and targeting. It typically includes steps such as defining objectives, gathering data, conducting analysis, segment creation, validation, and implementation planning. By implementing a structured process, organizations can enhance their understanding of target markets, develop tailored marketing strategies, allocate resources effectively, and improve overall decision-making. A structured process helps minimize bias, ensures consistency, and enables collaboration among different stakeholders involved in the segmentation process. It also facilitates documentation and evaluation of results, allowing for continuous improvement and refinement of segmentation strategies over time. Overall, implementing a structured process enhances the chances of successful market segmentation implementation and maximizes the benefits derived from the process.

Step 3: Collecting Data

This portion is contributed by – Joy Saha

1.Segmentation Variables

The book emphasizes the significance of empirical data in both commonsense and data-driven market segmentation. Empirical data forms the foundation for identifying and creating market segments and describing them in detail. In commonsense segmentation, a single characteristic, such as gender, is used as the segmentation variable to split the sample into segments. Other personal characteristics serve as descriptor variables, providing detailed information about the segments. This information helps develop an effective marketing mix for targeting each segment.

Data-driven market segmentation, on the other hand, relies on multiple segmentation variables. These variables, such as specific benefits sought by customers, are used to identify naturally existing or artificially created market segments. Descriptor variables, including socio-demographics and media behavior, further describe the segments. The quality of empirical data is crucial in developing valid segmentation solutions and accurately assigning individuals to the correct segments. It also enables the customization of products, determination of pricing and distribution strategies, and selection of effective communication channels.

The quality of data influences the effectiveness of both commonsense and data-driven segmentation approaches. Data can be obtained from various sources, including surveys, observations (such as scanner data), and experimental studies. While surveys are commonly used, their reliability in reflecting actual behavior, especially for socially desirable actions, may be limited. Therefore, it is essential to explore diverse data sources that closely align with consumer behavior when conducting segmentation studies. Ultimately, good market segmentation analysis relies on the availability of high-quality empirical data.

2. Segmentation Criteria

Before diving into market segmentation and data collection, organizations face a crucial decision of choosing the segmentation criterion. The term "segmentation criterion" refers to the nature of the information used for segmentation, encompassing various factors beyond a single measured value or variable. Common segmentation criteria include geographic, socio-demographic, psychographic, and behavioral dimensions.

The choice of segmentation criterion is not easily outsourced to consultants or data analysts, as it requires prior market knowledge. Several relevant differences among consumers, such as profitability, bargaining power, benefit preferences, barriers to choice, and consumer interaction effects, are highlighted by Bock and Uncles (2002) as key considerations for market segmentation.

With numerous segmentation criteria available, determining the best one to use can be challenging. Hoek et al. (1996) note the lack of clear guidelines regarding the most appropriate segmentation base for a given marketing context. In general, the recommendation is to adopt the simplest approach possible. Cahill (2006) succinctly advises, "Do the least you can." If demographic segmentation suits the product or service, it should be utilized. Likewise, if geographic segmentation aligns with regional appeal, it should be employed. The allure and sophistication of psychographic segmentation should not overshadow its effectiveness. Ultimately, the best segmentation criterion is the one that works for the product or service at the lowest cost.

Choosing the appropriate segmentation criterion requires careful consideration of the product, market, and objectives. It is important to focus on what resonates with the target audience and provides the most valuable insights, rather than getting caught up in complexity for the sake of sophistication.

2.1 Geographic Segmentation

Geographic segmentation, which uses the consumer's location of residence as the main criterion, has historically been the foundation of market segmentation. While straightforward, it is often a suitable approach. For example, the national tourism organization of Austria might use geographic segmentation to target tourists from neighboring countries due to language differences. Geographic segmentation allows for easy targeting of communication messages and the selection of relevant channels to reach specific geographic segments.

However, a limitation of geographic segmentation is that residing in the same area does not guarantee shared characteristics or preferences among consumers. Factors such as product preferences are often better explained by socio-demographic criteria rather than location alone. This is evident in tourism, where individuals from the same country may have diverse ideal holiday preferences based on personal interests and preferences.

Despite its limitations, geographic information has seen a resurgence in international market segmentation studies that aim to extract segments across geographic boundaries. However, conducting such studies poses challenges, including the need for meaningful segmentation variables applicable to diverse regions and accounting for cultural biases in survey responses.

Overall, geographic segmentation remains relevant and valuable in specific contexts, illustrating its enduring significance in market segmentation practices.

2.2 Socio-Demographic Segmentation

Socio-demographic segmentation criteria, such as age, gender, income, and education, are commonly used in various industries. They can be particularly relevant in sectors like luxury goods, cosmetics, baby products, retirement villages, and tourism resorts. Socio-demographic segmentation offers the advantage of easily determining segment membership for consumers. While certain product preferences can be attributed to socio-demographic factors (e.g., having children influencing vacation choices), in many cases, these criteria do not provide sufficient market insights for optimal segmentation decisions. Demographics alone explain only a small portion of consumer behavior, with values, tastes, and preferences being more influential factors. Yankelovich and Meer (2006) suggest that socio-demographics are not a strong basis for market segmentation compared to these other elements.

2.3 Psychographic Segmentation

Psychographic segmentation involves grouping individuals based on psychological criteria, such as beliefs, interests, preferences, aspirations, and sought benefits. It is a more complex approach compared to geographic or socio-demographic segmentation since it requires multiple variables to capture the psychographic dimension of interest. Benefit segmentation and lifestyle segmentation are popular forms of psychographic segmentation. The advantage of psychographic segmentation is that it provides insight into the underlying reasons for consumer behavior. For example, individuals motivated by cultural exploration are likely to prefer destinations with rich cultural offerings. However, determining segment memberships in psychographic segmentation can be challenging, and the effectiveness relies on reliable and valid measures of the psychographic dimensions being considered.

2.4 Behavioural Segmentation

Behavioral segmentation involves grouping individuals based on their behaviors or reported behaviors. This approach focuses on analyzing various aspects of behavior, such as prior product experience, purchase frequency, amount spent, and information search behavior. Behavioral segmentation has been found to be superior to geographic variables in segmenting tourists based on their reported behaviors. The key advantage of behavioral segmentation is that it directly uses the behavior of interest as the basis for segment extraction, ensuring relevance and accuracy. Actual behavioral data, rather than stated or intended behavior, provides a solid foundation for segmentation. It eliminates the need for developing measures of psychological constructs and allows for segmenting individuals based on their brand choices or purchase behavior over time. However, acquiring behavioral data may pose challenges, particularly when including potential customers who have not yet made a purchase.

3. Data from Survey Studies

Most market segmentation analyses are based on survey data. Survey data is cheap and easy to collect, making it a feasible approach for any organisation. But survey data – as opposed to data obtained from observing actual behaviour – can be contaminated by a wide range of biases. Such biases can, in turn, negatively affect the quality of solutions derived from market segmentation analysis.

3.1 Choice of Variables

The choice of variables is a critical factor in achieving high-quality market segmentation solutions. In data-driven segmentation, it is important to include all variables relevant to the segmentation criterion while avoiding unnecessary ones. Unnecessary variables can make surveys lengthy and tiresome for respondents, leading to lower response quality and respondent fatigue. Moreover, including irrelevant variables increases the complexity of segmentation analysis without providing valuable information, making it challenging to extract optimal market segments using data analytics techniques. Noisy variables, which do not contribute to identifying the correct market segments, can also hinder the segmentation process. They can arise from poorly designed survey questions or inadequate selection of segmentation variables. Such variables divert the attention of segmentation algorithms from extracting the accurate solution.

To address these issues, it is recommended to ask necessary and unique questions while refraining from including redundant or repetitive ones. Redundant questions, often resulting from traditional psychometric principles, can significantly interfere with the ability of segmentation algorithms to identify the correct market segmentation solutions. Developing an effective questionnaire typically involves conducting exploratory or qualitative research to gain insights into respondents' beliefs and preferences. These insights can then be incorporated into the questionnaire as answer options or variables, ensuring that critical variables are not overlooked. By employing a two-stage process involving both qualitative and quantitative research, the risk of omitting crucial variables can be minimized, ultimately improving the effectiveness of market segmentation.

3.2 Response Options

The response options provided to respondents in surveys play a crucial role in subsequent segmentation analysis. Different types of response options yield different scales of data, impacting the suitability of data analytic techniques for segmentation purposes.

Binary or dichotomous response options generate binary data, represented by 0s and 1s. The clear distance between these values poses no challenges for segmentation analysis. Nominal variables arise when respondents select a single option from a list of unordered categories, such as occupation. Nominal variables can be transformed into binary data by creating a binary variable for each answer option.

Metric data, obtained when respondents indicate a number (e.g., age or nights stayed), are ideal for segmentation analysis as they allow for statistical procedures and distance measurement. Ordinal data, derived from a limited number of ordered response options (e.g., Likert scales), present challenges due to the undefined distance between adjacent options. Specialized techniques or assumptions are needed to analyze such data accurately.

Ideally, binary or metric response options should be used in surveys to avoid complications related to distance measures during data-driven segmentation analysis. Binary options can outperform ordinal options, especially when formulated in a level-free manner. Visual analogue scales, like slider scales in online surveys, provide a metric approach by allowing respondents to indicate a position along a continuous line between two end-points.

Considering the appropriate response options based on the research question and the ability to capture nuances or response styles is essential in obtaining meaningful and effective data for segmentation analysis.

3.3 Response Styles

Survey data is prone to biases that can impact the reliability and interpretation of results. Response bias occurs when respondents consistently answer questionnaire items based on factors unrelated to the item content. These biases can introduce systematic errors and distort the outcomes of segmentation analysis.

Survey respondents exhibit various response styles, such as using extreme response options, consistently selecting the midpoint, or agreeing with all statements. These response styles can significantly influence segmentation results because commonly used algorithms struggle to differentiate between genuine beliefs and response biases. For example, an acquiescence bias, where respondents tend to agree with all questions, may create a segment that appears highly favorable but is actually influenced by the response style rather than actual preferences or behaviors.

To ensure accurate segmentation, it is crucial to minimize the risk of capturing response biases during data collection. When attractive market segments emerge with response patterns potentially driven by response biases, additional analyses are necessary to identify and address these biases. In some cases, it may be necessary to exclude respondents affected by response biases from the target market segment.

By recognizing and mitigating response biases, organizations can obtain more reliable and actionable insights from their market segmentation efforts, leading to better-informed decision-making.

3.4 Sample Size

The size of the sample used in market segmentation analysis plays a crucial role in obtaining accurate results. Inadequate sample sizes make it challenging to determine the correct number and nature of market segments. Several recommendations have been proposed. Formann suggests a sample size of at least 2p (or five times 2p), where p represents the number of segmentation variables. Qiu and Joe recommend a sample size of 10p times k, where k is the number of segments. Simulation studies by Dolnicar et al. indicate that a sample size of at least 60p is suitable for typical scenarios, while more challenging scenarios may require a sample size of 70p.

Market characteristics, such as segment size and overlap, and data characteristics, including biases, response styles, and correlation, influence the required sample size. It is important to collect unbiased and high-quality data with suitable response options to ensure reliable market segmentation analysis. A recommended guideline is to have a sample size of at least 100 respondents per segmentation variable to achieve accurate segmentation.

By having an adequate sample size and high-quality data, meaningful and actionable insights can be obtained from market segmentation analysis. It is crucial to consider these factors to ensure the validity and reliability of segmentation results and enable informed decision-making in marketing strategies and targeting specific market segments.

It can be concluded from the body of work studying the effects of survey data quality on the quality of market segmentation results based on such data that, optimally, data used in market segmentation analyses should

• contain all necessary items;

• contain no unnecessary items;

• contain no correlated items;

• contain high-quality responses;

• be binary or metric;

• be free of response styles;

• include responses from a suitable sample given the aim of the segmentation study;

• include a sufficient sample size given the number of segmentation variables (100 times the number of segmentation variables).

3.5 Data from Internal Sources

Organizations have access to abundant internal data for market segmentation analysis, including scanner data, booking data, and online purchase data. Internal data represents actual consumer behavior and eliminates biases associated with self-reported information. It is automatically generated and readily accessible. However, using only internal data may introduce systematic bias by over-representing existing customers and neglecting potential future customers. To ensure comprehensive segmentation, organizations should supplement internal data with external sources to capture a broader consumer base and make informed decisions.

3.6 Data from Experimental Studies

Experimental data, obtained from field or laboratory experiments, can be utilized for market segmentation analysis. These experiments can involve testing consumer responses to advertisements or conducting choice experiments and conjoint analyses. Through such studies, specific product attributes and their levels are presented to consumers, and their preferences are recorded. This information on consumer preferences and attribute effects can serve as valuable segmentation criteria for understanding consumer behaviour and making informed marketing decisions. Experimental data provides a controlled environment for studying consumer responses and can contribute to effective market segmentation strategies.

Step 4: Exploring Data

This article is taken from the report of – Manishankar Bag

A First Glimpse at the Data:

"A First Glimpse at the Data" is an initial step in data exploration where you take a quick overview of the dataset. You examine the basic properties of the data, such as the number of observations or records, the number of variables or features, and the data types of each variable. Understanding these characteristics helps you get a sense of the dataset's structure and assists in determining the appropriate analysis techniques to apply.

Data Cleaning:

Data Cleaning, also known as data cleansing or data scrubbing, is a critical step in the data exploration process. It involves identifying and addressing any inconsistencies, errors, or missing values present in the dataset. This ensures that the data is accurate and reliable for further analysis. Data cleaning tasks can include removing duplicate entries, imputing missing values using appropriate methods (such as mean imputation or regression imputation), handling outliers, correcting formatting issues, and resolving inconsistencies in variable values.

Descriptive Analysis:

Descriptive Analysis involves exploring and summarizing the characteristics of the dataset using various statistical measures and visualizations. It helps you understand the central tendencies (such as mean, median, and mode) and variabilities (such as standard deviation and range) of the variables in the dataset. Descriptive analysis also includes examining the distribution of data using techniques like histograms, box plots, and density plots. These descriptive statistics and visualizations provide insights into the dataset, such as the spread of data, the presence of outliers, and potential relationships between variables.

Pre-Processing:

Pre-processing refers to the preparation of the data before performing further analysis or modeling. It involves transforming and manipulating the data to ensure it is in a suitable format for the chosen analysis techniques. Pre-processing tasks may include feature scaling (such as normalization or standardization) to bring different variables to a common scale, handling missing values, encoding categorical variables into numerical representations (using techniques like one-hot encoding or label encoding), and splitting the dataset into training and testing subsets.

Categorical Variables:

Categorical variables are variables that represent qualitative or discrete characteristics or groups. Examples include gender (male/female), education level (high school/college/graduate), or product categories (electronics/clothing/furniture). Analyzing categorical variables involves summarizing the frequencies or proportions of different categories using techniques like frequency tables or bar charts. It helps you understand the distribution of data among different groups and can provide insights into patterns, associations, or dependencies between categorical variables.

Numeric Variables:

Numeric variables are variables that represent quantitative or numerical values. They can be continuous (such as height or weight) or discrete (such as age or number of siblings). Analyzing numeric variables involves calculating descriptive statistics, such as measures of central tendency (mean, median) and measures of variability (standard deviation, range). Visualizations like histograms, box plots, or scatter plots are commonly used to explore the distribution, spread, and relationships within numeric variables.

Principal Components Analysis (PCA):

Principal Components Analysis (PCA) is a popular dimensionality reduction technique used in data exploration and analysis. It aims to identify the underlying structure or patterns in a high-dimensional dataset by transforming the original variables into a new set of uncorrelated variables called principal components. PCA achieves this by projecting the data onto the directions of maximum variance. It helps in reducing the complexity of the dataset, visualizing high-dimensional data in lower dimensions, identifying important variables, and discovering latent features or patterns that may be hidden in the original data.

Overall, these topics play essential roles in the process of exploring data, understanding its properties, and preparing it for further analysis or modeling. They provide valuable insights into the dataset, help uncover relationships or patterns, and enable informed decision-making based on the data.

Step-5: Extracting Segments

This article is taken from the report of – Soumyajit Mishra amd Joy Saha

Grouping Consumers

Data-driven market segmentation analysis is exploratory by nature. Consumer data sets are typically not well structured. Consumers come in all shapes and forms; a two-dimensional plot of consumers' product preferences typically does not contain clear groups of consumers. Rather, consumer preferences are spread across the entire plot. The combination of exploratory methods and unstructured consumer data means that results from any method used to extract market segments from such data will strongly depend on the assumptions made on the structure of the segments implied by the method. The result of a market seg-mentation analysis, therefore, is determined as much by the underlying data as it is by the extraction algorithm chosen. Segmentation methods shape the segmentation solution. Many segmentation methods used to extract market segments are taken from the field of cluster analysis. In that case, market segments correspond to clusters. As pointed out by Hennig and Liao (2013), selecting a suitable clustering method requires matching the data analytic features of the resulting clustering with the context-dependent requirements that are desired by the researcher(p. 315). It is, therefore, important to explore market segmentation solutions derived from a range of different clustering methods. It is also important to understand how different algorithms impose structure on the extracted segments. If consumer data is well-structured, and well-separated, distinct market segments exist, tendencies of different algorithms matter less. If, however, data is not well-structured, the tendency of the algorithm influences the solution substantially. In such situations, the algorithm will impose a structure that suits the objective function of the algorithm. So called distance-based methods are described first. Distance-based methods use a particular notion of similarity or distance between observations (consumers), and try to find groups of similar observations (market segments). So called model- based methods are described second. These methods formulate1 a concise stochastic model for the market segments. In addition to those main two groups of extraction methods, a number of methods exist which try to achieve multiple aims in one step. For example, some methods perform variable selection during the extraction of market segments. A few such specialized algorithms are also discussed. Because no single best algorithm exists, investigating and comparing alternative segmentation solutions is critical to arriving at a good final solution. Data char act eristics and expected or desired segment characteristics allow a pre-selection of suitable algorithms to be included in the comparison.

5.1 Distance-Based Methods

1.2 Distance-Based Methods

Consider the problem of finding groups of tourists with similar activity patterns when on vacation. A fictitious data set is shown in Table 1. It contains seven people indicating the percentage of time they spend enjoying BEACH, ACTION, and CULTURE when on vacation. Anna and Bill only want to relax on the beach, Frank likes beach and action, Julia and Maria like beach and culture, Michael wants action and a little bit of culture, and Tom does everything. Market segmentation aims at grouping consumers into groups with similar needs or behavior, in this example: groups of tourists with similar patterns of vacation activities. Anna and Bill have exactly the same profile, and should be in the same segment. Michael is the only one not interested in going to the beach, which differentiates him from the other tourists. In order to find groups of similar tourists one needs a notion of similarity or dissimilarity, mathematically speaking: a distance measure.

1.2.1 Distance Measures

Table 1 illustrates a data matrix denoted as X, where each row represents an observation (tourist) and each column represents a variable (vacation activity). Mathematically, X can be represented as an n × p matrix, where n denotes the number of observations (rows) and p denotes the number of variables (columns).

Euclidean distance is the most commonly used measure in market segmentation analysis. It represents the direct ”straight-line” distance between two
points in two-dimensional space. On the other hand, Manhattan distance de-
rives its name from considering the distance between two points using grid-like
structures, such as the street layout in Manhattan. Both Euclidean and Manhattan distances utilize all dimensions of the vectors x and y. The asymmetric binary distance does not use all dimensions of the vectors. It only uses dimensions where at least one of the two vectors has a value of 1. It is asymmetric because it treats 0s and 1s differently. Similarity between two observations is only concluded if they share 1s, but not if they share 0s. The dissimilarity between two observations is increased if one has a 1 and the other not. This has implications for market segmentation analysis. The asymmetric binary distance corresponds to the proportion of common 1s over all dimensions where at least one vector contains a 1. A symmetric binary distance measure emerges from using the Manhattan distance between the two vectors, treating 0s and 1s equally. Euclidean distance is the most commonly used measure in market segmentation analysis. It represents the direct ”straight-line” distance between two points in two-dimensional space. On the other hand, Manhattan distance derives its name from considering the distance between two points using grid-like structures, such as the street layout in Manhattan. Both Euclidean and Manhattan distances utilize all dimensions of the vectors x and y.The asymmetric binary distance does not use all dimensions of the vectors. It only uses dimensions where at least one of the two vectors has a value of 1. It is asymmetric because it treats 0s and 1s differently. Similarity between two observations is only concluded if they share 1s, but not if they share 0s. The dissimilarity between two observations is increased if one has a 1 and the other not. This has implications for market segmentation analysis. The asymmetric binary distance corresponds to the proportion of common 1s over all dimensions where at least one vector contains a 1. A symmetric binary distance measure emerges from using the Manhattan distance between the two vectors, treating 0s and 1s equally.

5.1.1 Hierarchical Methods

Hierarchical clustering is a method for organizing data into a hierarchy of clusters. It starts by considering each data point as an individual cluster and then iteratively combines similar clusters to form larger clusters or splits clusters to create smaller ones. The result is a hierarchical structure, often represented as a dendrogram, which illustrates the relationships between clusters at different levels of granularity.

There are two main types of hierarchical clustering: agglomerative and divisive.

1. Agglomerative hierarchical clustering: It begins with each data point as a separate cluster and merges the closest pair of clusters iteratively, based on a distance or similarity measure. The process continues until all data points belong to a single cluster. Agglomerative clustering can be visualized as a bottom-up approach, where smaller clusters are successively merged to form larger clusters.

2. Divisive hierarchical clustering: It starts with all data points belonging to a single cluster and then divides the clusters recursively into smaller subclusters. At each step, the algorithm selects a cluster and splits it into two based on a dissimilarity criterion. This process continues until each data point is in its own individual cluster. Divisive clustering can be visualized as a top-down approach, where larger clusters are successively divided into smaller clusters.

5.1.2 Partitioning Methods

Hierarchical clustering methods are often used for the analysis of small data sets with 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, partitioning clustering methods are more suitable. These methods aim to create a single partition of the data into segments rather than a nested sequence of partitions. One popular partitioning method is k-means clustering, which divides the data into subsets or segments based on their similarity.

5.1.2.1 k-Means and k-Centroid Clustering

k-means clustering is the most commonly used distance based partitioning clustering algorithm. Using random consumers from the data sets 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. Generalisations using other distances are also referred to as k-centroid clustering.

The k-means algorithm involves several steps:

1.Specify the desired number of segments, denoted as k.

2.Randomly select k observations as initial cluster centroids.

3.Assign each observation to the closest cluster centroid to form an initial partition.

4.Recompute the cluster centroids based on the current partition.

5.Repeat steps 3 and 4 until convergence or a maximum number of iterations is reached.

5.1.2.2 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 smart starting values instead of randomly selecting consumers from the data set. The best starting points are 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.

5.1.2.3 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. This procedural difference can lead to different segmentation solutions compared to k-means, and it may find the globally optimal solution while k-means gets stuck in a local optimum.

5.1.2.4 Neural Gas and Topology Representing Networks:

Neural gas and topology representing networks are further extensions of hard competitive learning. Neural gas adjusts not only the closest representative but also the second closest representative towards the randomly selected consumer, with a smaller adjustment for the second closest. Topology representing networks count 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.

5.1.2.5 Self-organizing maps:

Self-organizing maps are another variation of hard competitive learning that positions segment representatives on a regular grid. It use a rectangular or hexagonal grid and adjust 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.

5.1.2.6 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 segment representatives, while the parameters connecting the input layer to the hidden layer indicate membership in different segments.

5.1.3 Hybrid Approaches:

Hybrid segmentation approaches aim to leverage the strengths 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.

5.1.3.1 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 its 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 centers 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.

5.1.3.2 Bagged Clustering:

Bagged clustering, on the other hand, combines hierarchical clustering and partitioning clustering with bootstrapping. It starts by creating multiple bootstrap samples from the original data set. 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.

5.2 Model-Based Methods:

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

5.2.1 Finite Mixtures of 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 membership are calculated based on consumer information and the estimated parameter values. The segment with the highest probability is then assigned to the consumer.

5.2.1.1 Normal Distributions:

A mixture of normal distributions 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 modeling covariance between variables, and it commonly occurs in biological and business contexts. For instance, physical measurements on humans, such as height, arm length, leg length, or 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 μ_h of length p. In addition to the variances of the p segmentation variables, the covariance structure can also be modeled. This results in a $p \times p$ covariance matrix Σ_h for each segment. The covariance matrix Σ_h contains the variances of the p segmentation variables on the diagonal and the covariances between pairs of segmentation variables in the other entries. The covariance matrix is symmetric and contains $p(p + 1)/2$ unique values.

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

5.2.1.2 Binary Distributions:

The finite mixtures of binary distributions to model market segmentation based on binary segmentation variables representing customer 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 segment-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, the mixture of binary distributions provides a way to model the association between binary variables and identify distinct market segments based on activity patterns.

5.2.2 Finite Mixtures of Regressions:

Finite mixtures of regressions provide a different perspective on market segmentation compared to distance-based clustering methods. They analyze the relationship between variables and allow for the identification of distinct segments with varying regression patterns. Finite mixture of regression models assume 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 different for different market segments.

5.3 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 selection plays a crucial role in segmentation analysis, especially when dealing with redundant or noisy variables. Different algorithms, such as biclustering, VSBD, and factor-cluster analysis, offer integrated approaches to segment extraction while simultaneously selecting suitable segmentation variables, taking into account the specific characteristics of binary data.

5.3.1 Biclustering Algorithms:

Biclustering is a method for simultaneously clustering consumers and variables in market segmentation analysis. Biclustering algorithms 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 a large number of 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 segmentation analysis, particularly with binary data and a large number of variables. It enables the identification of groups of consumers and variables with common patterns, providing insights into niche markets and avoiding data transformation biases.

5.3.2 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.

5.3.3 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.

5.4 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:

5.3.1 Cluster indices:

Cluster indices provide measures 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.

5.3.2 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.

5.3.3 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.

5.3.4 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.

Step 6: Profiling Segments

This article is taken from the report of – Manish Kumar

6.1 Identifying Key Characteristics of Market Segments

The profiling step in market segmentation is essential for understanding the resulting market segments from the extraction step, particularly in data-driven segmentation. Profiling involves identifying the defining characteristics of each market segment with respect to the segmentation variables. Unlike commonsense segmentation, where the segment profiles are predefined (e.g., age groups), data-driven segmentation requires the analysis of data to uncover the characteristics of the segments.

Profiling aims to characterize each market segment individually and compare them to other segments. It helps differentiate segments based on their unique characteristics. For example, if winter tourists in Austria are surveyed about their vacation activities, most may mention alpine skiing. While alpine skiing can be a characteristic of a segment, it may not be sufficient to differentiate that segment from others. Therefore, profiling is crucial for a thorough understanding of the segments and for making effective strategic marketing decisions.

However, data-driven market segmentation solutions can be challenging to interpret. Many managers struggle with understanding the results and view segmentation analysis as a black box. They often receive lengthy reports that may contradict the results or lack clear executive summaries. The presentation of segmentation results can be rushed and confusing, typically presented in numbers and percentages across several variables, leaving managers with insufficiently conclusive information. To address these challenges, graphical statistics approaches are recommended as they make profiling less tedious and prone to misinterpretation.

In summary, profiling plays a vital role in data-driven market segmentation by identifying the defining characteristics of each segment. It helps marketers make informed decisions by understanding the unique attributes of each segment. However, interpreting segmentation results can be challenging, and graphical statistics approaches offer a more accessible and less ambiguous way to profile market segments.

6.2 Traditional Approaches To Profiling Market Segments

We use Australian vacation motives dataset and the extraction of segments using the neural gas clustering algorithm. The resulting segmentation solution is presented and the challenges of interpreting data-driven segmentation solutions are highlighted.

Data-driven segmentation solutions are often presented in two ways: oversimplified summaries or large tables that are difficult to interpret. These tables provide mean values or percentages of segmentation variables for each segment. Interpreting the defining characteristics of each segment requires comparing these values with those of other segments or the total.

For example, segment 2 in the presented solution is characterized by a preference for rest and relaxation, staying within the planned travel budget, and a lesser emphasis on cultural offers, intense nature experiences, prices, health and beauty, and creativity. On the other hand, segment 1 appears to be a response style segment with relatively low percentages indicating relevance for each travel motive compared to the overall agreement percentage.

Interpreting all six market segments based on the provided segmentation solution requires a significant number of comparisons, both between segments and with the total. This can be a tedious and challenging task, especially when considering multiple alternative segmentation solutions.

In some cases, information about the statistical significance of differences between segments may be provided. However, using standard statistical tests to assess significance is not appropriate due to the nature of segment creation and maximization of differences.

In summary, interpreting data-driven segmentation solutions can be challenging due to the extensive number of comparisons involved. Simplified summaries or alternative approaches are necessary to facilitate a clearer understanding of the defining characteristics of each market segment.

6.3 Segment Profiling With Visualizations

The traditional tabular representations commonly used to present market segmentation solutions often overlook the potential of graphics, despite the integral role of data visualization in statistical data analysis. Graphics have significant advantages in exploratory statistical analysis, such as cluster analysis, as they provide insights into complex variable relationships. Moreover, in the era of big data, visualization offers a straightforward means of monitoring temporal developments.

Various scholars, including McDonald and Dunbar (2012) and Lilien and Rangaswamy (2003), recommend the utilization of visualization techniques to enhance the interpretation of market segmentation analysis results. Even earlier, Haley (1985) emphasized that tabular presentations lack the same level of insightfulness as graphical representations. Recent research by Cornelius et al. (2010) further supports the notion that simpler two-dimensional graphical formats are preferable due to their intuitive interpretations compared to more complex representations.

Numerous examples demonstrate the use of visualizations in interpreting segmentation solutions, as discussed in studies by Reinartz and Kumar (2000), Horneman et al. (2002), Andriotis and Vaughan (2003), Becken et al. (2003), Dolnicar and Leisch (2003, 2014), Bodapati and Gupta (2004), Dolnicar (2004), Beh and Bruyere (2007), and Castro et al. (2007).

In the data-driven market segmentation process, visualizations play a crucial role in examining and interpreting segment profiles for each segmentation solution. Statistical graphs aid in understanding the characteristics of segments and facilitate the assessment of the usefulness of a particular market segmentation solution. Given the numerous alternative solutions that arise during the segmentation process, selecting the most appropriate solution becomes a critical decision, where visualizations provide valuable support to data analysts and users.

6.4 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.
- Visualizations, 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 make the interpretation of 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 color 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 interpretation and comprehension of segmentation results. Well-designed graphs offer a valuable return on investment, especially for managers making strategic decisions based on segmentation outcomes.

6.5 Assessing Segment Separation:

Segment separation can be evaluated using segment separation plots, which provide a visual representation of the overlap among segments in the data space. These plots offer a quick overview of the data situation and the segmentation solution, making it easier for data analysts and users to assess segment separation.

In segment separation plots, observations are depicted in scatter plots, with each observation colored according to its segment membership. The cluster hulls, representing the shape and spread of the true segments, are also displayed. Dashed cluster hulls typically contain all or most observations, while solid cluster hulls represent a subset of observations. Neighbourhood graphs, indicated by black lines with numbered nodes, illustrate the similarity between segments.

For lower-dimensional data, such as two-dimensional data, segment separation plots can be directly plotted without the need for projection. However, for high-dimensional data, projection techniques like principal components analysis may be employed to reduce the dimensions and create a segment separation plot.

The resulting segment separation plot provides a visual representation of the segmentation solution, with each segment's center indicated by a numbered node. The black lines connecting segment centers signify the similarity between segments, with thicker lines indicating more observations sharing those segment centers as their closest.

It's important to note that segment separation plots depict one possible projection, and overlapping segments in a particular projection do not imply overlap in all projections. Careful interpretation of the plot is necessary to understand the distinct characteristics of each segment and their travel motives.

By employing modifications like color adjustments, omitting observations, and highlighting specific areas of segments, a cleaner and more interpretable version of the segment separation plot can be created. This enhanced plot combines the advantages of the segment separation plot with perceptual maps, facilitating a better understanding of the market segments.

In summary, segment separation plots are valuable tools for assessing segment separation in market segmentation analysis. They provide visual representations of segment overlap and help identify distinct characteristics of each segment. By utilizing appropriate projection techniques and enhancing the plots, analysts can gain insights into the separation between market segments and their specific travel motives.

Step 7: Describing Segments

This article is taken from the report of – Sudip Sahoo

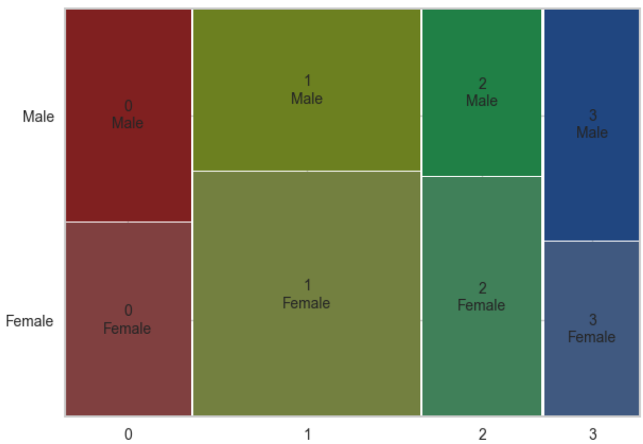
7.1 Developing a Complete Picture of Market Segments

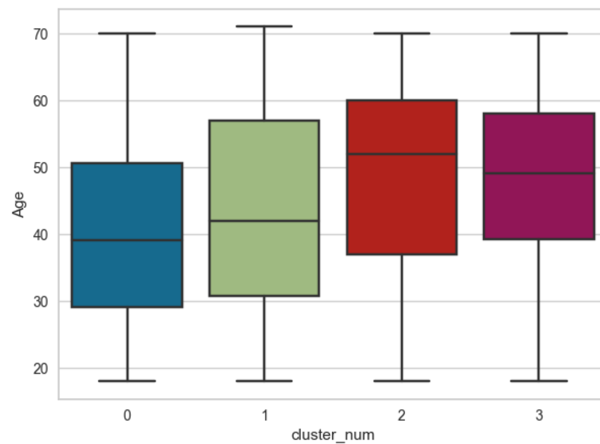
Understanding the variations in segmentation factors across market segments is the goal of segment profiling. The selection of segmentation variables occurs conceptually in Step 2 (definition of the ideal target segment) and empirically in Step 3 (data collection).

The foundation for deriving market segments from empirical data are segmentation variables. Comparable to the profiling stage is step 7 (describing segments). The sole distinction is that market segmentation has not been done using the characteristics under examination. Instead, additional information about segment participants is used to describe market segments in Step 7.

If choosing a target segment is like getting married, then profiling and characterising market segments is like going on several dates to get to know the potential spouse as much as you can in an effort to give the marriage the greatest opportunity and prevent unpleasant shocks down the road.

For example, when conducting a data-driven market segmentation analysis on McDonald's food data set. The fast-food data set is not typical for data collected for market segmentation analysis because it contains very few descriptor variables. Descriptor variables – additional pieces of information about consumers – are critically important to gaining a good understanding of market segments. One descriptor variable available in the fast-food data set is the extent to which consumers love or hate McDonald's. Using a simple mosaic plot, we can visualise the association between segment membership and loving or hating McDonald's.





To do this, we first extract the segment membership for each consumer for the four-segment solution. Next, we cross-tabulate segment membership and the love-hate variable. Finally, we generate the mosaic plot with cells colours indicating the deviation of the observed frequencies in each cell from the expected frequency if variables are not associated.

7.2 Using Visualizations to Describe Market Segments

There are numerous charts available for displaying the differences between descriptor variables. Here, we go over two fundamental methods that can be used with nominal and ordinal descriptor variables (like gender, education level, or country of origin) or metric descriptor variables (like age, the number of nights spent travelling, or the amount of money spent on lodging). Two major benefits of using graphical statistics to describe market segments are that it makes it easier for users and data analysts to understand results and that it incorporates information on the statistical significance of differences, preventing the over-interpretation of insignificant differences.

7.2.1 Nominal and Ordinal Descriptor Variables

A cross-tabulation of segment membership with the descriptor variable serves as the foundation for all visualisations and statistical tests when differentiating between market segments using a single nominal or ordinal descriptor variable.

Segment number is plotted along the x-axis of the mosaic plot in below figure, and liking or disliking McDonald's is plotted along the y-axis. A strong and significant correlation between those two variables is shown by the mosaic plot. The top left boxes are red to show that members of segment 1 (seen in the first column) rarely express fondness for McDonald's. The dark blue boxes in the top right corner of the mosaic plot show that people in segment 4 are noticeably more likely to adore McDonald's. These customers are also less likely to despise McDonald's, as shown by the tiny red boxes in the bottom right corner of the plot.

Additionally, mosaic plots can incorporate components of inferential statistics and show tables with more than two descriptor variables. This makes interpretation easier. Cellular colours can be used to indicate areas were observed and expected frequencies diverge, if the variables are assumed to be independent. Based on the standardised difference between the expected and observed frequencies, cell colours are determined.

The fast-food data contains a few other basic descriptor variables, such as gender and age. Figure A.12 shows gender distribution across segments.

7.2.2 Metric Descriptor Variables

Conditional in this context refers to the division of the plots into sections (panels, facets), each of which presents the outcomes for a certain subset of the data (for instance, various market segments). Using metric descriptor variables, conditional charts are a good tool for illustrating variations between market segments.

7.3 Testing for Segment Differences in Descriptor Variables

Formally testing for variations in descriptor variables across market groups can be done using straightforward statistical tests. Running a set of separate tests for each relevant variable is the simplest way to look for differences. Segment membership, or the assignment of each consumer to a certain market segment, is the result of the segment extraction stage. Segment membership is a nominal variable that can be handled just like any other. It serves as the segmentation variables' nominal summary statistic. Therefore, any test to determine if a nominal variable is associated with another variable is appropriate.

The appropriate test for independence between columns and rows of a table is the $\chi2$-test. The output contains: the name of the statistical test, the data used, the value of the test statistic (in this case X-squared), the parameters of the distribution used to calculate the p-value (in this case the degrees of freedom (df) of the $\chi2$- distribution), and the p-value. The p-value indicates how likely the observed frequencies occur if there is no association between the two variables (and sample size, segment sizes, and overall gender distribution are fixed). Small p-values (typically smaller than 0.05), are taken as statistical evidence of differences in the gender distribution between segments. Here, if test results in a non-significant p-value, implying that the null hypothesis is not rejected.

If the $\chi2$-test rejects the null hypothesis of independence because the p-value is smaller than 0.05, a mosaic plot is the easiest way of identifying the reason for rejection. The colour of the cells points to combinations occurring more or less frequently than expected under independence. The association between segment membership and metric variables (such as age, number of nights at the tourist destinations, dollars spent on accommodation) is visualised using parallel boxplots. Any test for difference between the location (mean, median) of multiple market segments can assess if the observed differences in location are statistically significant.

The most popular method for testing for significant differences in the means of more than two groups is Analysis of Variance (ANOVA).

7.4 Predicting Segments from Descriptor Variables

Another way of learning about market segments is to try to predict segment membership from descriptor variables. To achieve this, we use a regression model with the segment membership as categorical dependent variable, and descriptor variables as independent variables. We can use methods developed in statistics for classification, and methods developed in machine learning for supervised learning.

As opposed to the methods in Sect. 9.3, these approaches test differences in all descriptor variables simultaneously. The prediction performance indicates how well members of a market segment can be identified given the descriptor variables. We also learn which descriptor variables are critical to the identification of segment membership, especially if methods are used that simultaneously select 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 $x_1, x_2, ..., x_n$:

$$y \approx f\left(x_1, x_2, ..., x_n\right)$$

Regression models differ with respect to the function $f(.)$, the distribution assumed for y, and the deviations between y and $f\left(x_1, x_2, ..., x_n\right)$. The basic regression model is the linear regression model. The linear regression model assumes that function $f(.)$ is linear, and that y follows a normal distribution with mean $f\left(x_1, x_2, ..., x_n\right)$ and variance σ^2. The relationship between the dependent variable y and the independent variables $x_1, x_2, ..., x_n$ is given by:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$$

where $N\left(\mu, \sigma^2\right)$.

Regression coefficients in linear regression models describe how much the dependent variable shifts when one independent variable shift while the other independent variables stay the same. According to the linear regression model, changes brought about by adjustments to one independent variable are independent of changes in the absolute levels of the other independent variables.

In the linear regression model, the dependent variable has a normal distribution. A greater variety of distributions for the dependent variable can be accommodated by generalised linear models (Nelder and Wedderburn 1972). This is crucial if the dependent variable has a categorical nature and the normal distribution is thus inappropriate.

In the linear regression model, the mean value of y given $x_1, x_2, ..., x_n$ is modelled by the linear function:

$$E\left[y \mid x_1, x_2, ..., x_p\right] = \mu = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

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 describe the mean value with a linear function which can take any real value. Generalised linear models account for this by introducing a link function $g(.)$. The link function transforms the mean value of y given by μ to an unlimited range indicated by η. This transformed value can then be modelled with a linear function:

$$g(\mu) = \eta = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

Here η is referred as linear predicator.

7.4.1 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 μ, and by choosing the logit link that maps the success probability $\mu \in (0,1)$ onto $(-\infty, \infty)$ by

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

The intercept in the linear regression model gives the mean value of the dependent variable if the independent variables $x_1, x_2, ..., x_n$ all have a value of 0. In binomial logistic regression, the intercept gives the value of the linear predictor η if the independent variables $x_1, x_2, ..., x_n$ all have a value of 0.

The other regression coefficients in a linear regression model indicate how much the mean value of the dependent variable changes if this independent variable changes while others remain unchanged. In binary logistic regression, the regression coefficients indicate how the linear predictor changes. The changes in the linear predictor correspond to changes in the log odds of success. The odds of success are the ratio between the probability of success μ and the probability of failure $1 - \mu$. 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.

7.4.2 Multinomial Logistic Regression

The multinomial logistic regression method can fit a model that simultaneously predicts each segment. The dependent variable y is not binary because segment extraction often yields more than two market segments. Instead, it is assumed to be categorical and to follow a multinomial distribution with the logistic function acting as the link function.

Advantages:

- Helps to understand the relationships among the variables present in the dataset.
- Simultaneous Models result in smaller standard errors for the parameter estimates than when fitting the logistic regression models separately.
- The choice of reference class has no effect on the parameter estimates for other categories.

7.4.3 Tree-Based Methods

An alternative modelling strategy for predicting a binary or categorical dependent variable given a set of independent factors is classification and regression trees (CARTs; Breiman et al. 1984). Trees used for classification and regression are an example of supervised learning in machine learning. The benefits of classification and regression trees include their capacity to execute variable selection, straightforward integration of interaction effects, and ease of interpretation aided by visuals. Regression and classification trees perform effectively when there are many independent variables. The drawback is that outcomes are frequently erratic. Completely distinct trees can result from even minor modifications in the data.

The model is fitted using a stepwise process in the tree approach. Consumers are divided into groups based on one independent variable at each phase. The goal of the split is to produce groups that are as homogeneous as feasible in terms of the dependent variable. This indicates that the dependent variable's values are comparable among consumers in the resulting groupings. In the ideal scenario, a categorical dependent variable's value is the same for every member of the group. The classification and regression tree approach is also known as recursive partitioning because of this sequential splitting process.

The resulting tree that has the nodes that emerge from each splitting step. The node containing all consumers is the root node. Nodes that are not split further are terminal nodes. We predict segment membership by moving down the tree. At each node, we move down the branch reflecting the consumer's independent variable. When we reach the terminal node, segment membership can be predicted based on the segment memberships of consumers contained in the terminal node. Tree constructing algorithms differ with respect to:

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

Information on product preferences, how often people eat at fast food establishments or out at restaurants generally, their hobbies, and how often they use media outlets (including TV, radio, newspapers, and social media) for information would all be of great interest. Due to the availability of this data, the data analyst may create a thorough description of each market category. A thorough description then forms the basis for the activities carried out in Step 9, where the ideal marketing mix for the chosen target segment is designed.

Step 8: Selecting the Target Segment(s)

This article is taken from the report of – Manishankar Bag

8.1 Selecting the Target Segment(s):

Selecting the target segment(s) refers to the process of identifying specific groups of customers or market segments that a business intends to focus its marketing efforts on. This involves analyzing and evaluating different segments based on criteria such as demographics, psychographics, behaviors, or needs. By selecting target segments, businesses can allocate their resources effectively and tailor their marketing strategies to meet the specific preferences and demands of those segments.

8.2 The Targeting Decision:

The targeting decision involves choosing the target segment(s) that a business will prioritize for its marketing efforts. It requires considering factors such as the size and growth potential of the segment, the competition within the segment, and the fit between the business's offerings and the segment's needs. The targeting decision is crucial as it determines the direction and scope of the marketing strategy and influences the allocation of resources, messaging, and positioning of the business's products or services.

8.3 Market Segment Evaluation:

Market segment evaluation involves assessing the attractiveness and viability of different market segments. This evaluation helps businesses understand the potential value and profitability of each segment. Factors considered during market segment evaluation may include the segment's size, growth rate, profitability, competition intensity, customer loyalty, and compatibility with the business's capabilities and resources. Market segment evaluation enables businesses to prioritize and invest in segments that offer the highest potential for success and align with their strategic goals.

In summary, selecting the target segment(s) involves identifying specific groups of customers or market segments to focus on, the targeting decision involves choosing the priority segments for marketing efforts, and market segment evaluation entails assessing the attractiveness and viability of different segments. These steps collectively contribute to developing effective marketing strategies, allocating resources efficiently, and maximizing the business's chances of success in the target market.

Step 9: Customizing the Marketing Mix

This article is taken from the report of – Manish Kumar

9.1 Implications for Marketing Mix Decisions

The marketing mix plays a crucial role in achieving successful sales results by combining various elements. Initially, marketing was seen as a toolbox for selling products, with marketers utilizing different ingredients to maximize outcomes. One of the popular frameworks is the 4Ps model, comprising Product, Price, Promotion, and Place. However, effective market segmentation should not be viewed in isolation but rather integrated with other strategic marketing aspects such as positioning and competition. The segmentation-targeting-positioning (STP) approach emphasizes a sequential process, starting with market segmentation, followed by targeting and positioning. Nonetheless, it's important not to strictly adhere to this sequence, as adjustments may be necessary before committing to specific target segments. Customizing the marketing mix to the chosen target segment is crucial for maximizing the benefits of a segmentation strategy. This involves product design, pricing modifications, suitable distribution channels, and appealing communication messages. The choice of segmentation variables depends on the purpose of the analysis, such as pricing, advertising, or distribution decisions. Ultimately, the insights gained from the target segment description guide the organization in developing or adjusting the marketing mix to cater to the chosen target segment effectively.

2.Product

When developing the product dimension of the marketing mix, organizations must consider customer needs. This may involve modifying an existing product rather than creating a completely new one. Other product-related decisions include naming the product, packaging, offering warranties, and providing after-sales support services. Target segment selection plays a crucial role in product design or modification. For instance, if targeting a segment with a strong interest in cultural activities, such as visiting museums, monuments, and gardens, the organization could develop a product like "MUSEUMS, MONUMENTS & MUCH, MUCH MORE" accompanied by an activities pass. This product would assist segment members in finding relevant activities and highlight them during the vacation planning process. Another opportunity is to promote gardens at the destination as an attraction in their own right, catering specifically to this segment's preferences. By aligning the product with the needs and interests of the target segment, organizations can effectively meet customer demands and enhance their marketing strategy.

3.Price

When developing the price dimension of the marketing mix, organizations must make key decisions regarding product pricing and discounts. For instance, in the context of a tourist destination targeting segment 3, identified through biclustering analysis of the Australian vacation activities data set, insights can be derived. By creating a segment membership vector, it becomes evident that segment 3 comprises a substantial number of consumers compared to other segments.

Further analysis reveals that segment 3 members exhibit higher expenditures per person per day during their vacations compared to other tourists. This presents a favorable opportunity for the destination as it suggests that they can set prices for the "MUSEUMS, MONUMENTS & MUCH, MUCH MORE" product without needing to offer discounts. In fact, there may even be potential to attach a premium price to this product, considering the segment's willingness to spend more.

Although more comprehensive information on price elasticity and consumer willingness to pay would be valuable, the obtained insights from this analysis highlight the strategic importance of the price dimension. By understanding the spending patterns and preferences of the target segment, organizations can effectively optimize their pricing strategies to maximize the impact of their targeted marketing efforts.

Overall, the price dimension plays a vital role in marketing decision-making. It requires careful consideration and alignment with the needs and behaviors of the target segment. By setting appropriate prices and utilizing discounts strategically, organizations can capitalize on market opportunities and enhance their competitiveness in attracting and satisfying their target customers.

4.Place

The place dimension of the marketing mix involves determining how the product should be distributed to customers. This decision encompasses factors such as online versus offline availability, direct sales versus intermediaries, and the use of wholesalers or retailers. In the context of segment 3 and a destination with a rich cultural heritage, insights from the market segmentation analysis provide valuable information on booking preferences.

By analyzing the survey responses on accommodation booking during respondents' last domestic holidays, the destination can align the distribution channels with the preferences of segment 3 members. Visualizing the booking behavior using tools like propBarchart allows for a clearer understanding. The resulting plot reveals that segment 3 members have a higher propensity for online hotel bookings compared to the average tourist.

This information has significant implications for the place dimension of the marketing mix. To effectively cater to segment 3, the destination must ensure the availability of an online booking option for hotels. Additionally, it would be advantageous to gather data on the booking behaviors of segment 3 members for other products, services, and activities to assess the extent of their online booking tendencies beyond accommodations.

By aligning the distribution strategy with the preferences and behaviors of the target segment, organizations can optimize the place dimension of the marketing mix. Providing the right distribution channels, whether online or offline, and making the product easily accessible to customers contribute to enhancing the overall customer experience and increasing the chances of reaching the target market effectively.

5.Promotion

Promotion decisions play a crucial role in the marketing mix and involve developing an advertising message that resonates with the target market and determining the most effective communication channels. Other promotional tools include public relations, personal selling, and sponsorship. In the context of segment 3, it is important to identify the best information sources for reaching these individuals and informing them about the MUSEUMS, MONUMENTS & MUCH, MUCH MORE product.

To determine the preferred information sources, an analysis of the information channels used by segment 3 members for their last domestic holiday is conducted. Using the propBarchart function, a plot is generated comparing the use of different information sources, such as tourist centers, in the decision-making process. The plot reveals that segment 3 members rely more heavily on information provided by tourist centers compared to other tourists. This insight can be utilized in designing the promotion component of the marketing mix by ensuring the availability of specific information packs about the product both in physical form at local tourist information centers and online on the tourist information center's website.

Furthermore, the TV channel preferences of segment 3 members are examined using a mosaic plot. The plot demonstrates that members of segment 3 have a preference for Channel 7, distinguishing them from other tourists. This information is valuable for developing a media plan that maximizes exposure to the targeted communication of the MUSEUMS, MONUMENTS & MUCH, MUCH MORE product, ensuring that the advertising message reaches the intended audience effectively.

By understanding the preferred information sources and media preferences of the target segment, organizations can tailor their promotional strategies to effectively communicate the value proposition of their products or services. This enables them to engage with their target market in a way that resonates with their preferences and increases the likelihood of capturing their attention and generating interest.

References

[1] S. Dolnicar, B. Grün and F. Leisch, "*Market Segmentation Analysis : Understanding It, Doing It, and Making It Useful*," Springer Singapore, 2018.