Market Segmentation Analysis Report

Team :

Nisarg Patel
Gaurav Deore
Harsh Priyam
Aradhya Kanth
Sania Shaikh

Introduction

The purpose of marketing is to match the genuine needs and desires of consumers with the offers of suppliers which particularly suit their needs.

A marketing plan consists of two components :
1. Strategic marketing plan
2. Tactical marketing plan

A strategic marketing plan is a which states the long term goal of the company but does not provide the short actions which will lead to achieving that goal , these short term actions or instructions on what needs to be done to get to the long term goal are stated in the tactical plan.

In most cases the tactical decisions depend on strategy listed in the strategic plan.

The Strategic Marketing Plan

It typically identifies consumer needs and desires , internal strengths and weaknesses of the organisation and external threats and opportunities that the organisation has to keep in mind.

Two methods
- SWOT analysis
- Market research (for consumer needs and desires)
  Market research has a heavy reliance on survey methodology used.

After the above analysis two key decisions have to be made as a part of the strategic marketing planning process.

- Target audience .. which consumers to focus on (segmentation and targeting)
- Branding … Which image of the organization should the consumers hold?

These decisions are critical because they determine the long-term direction of the organisation, and cannot easily be reversed

The Tactical Marketing Plan (usually period of upto 1 year)

Major four areas to focus on :-
- The development and modification of the product in view of needs and desire of target segment. (product)
- Determination of the price in view of cost ,competition and the willingness to pay of the target segment (Price)

- Selection of the most suitable distribution channels to reach the target segment (Place)
- Communication and promotion of the offer in a way that is most appealing to the target segment (Promotion)

As long as the strategic marketing is good the expedition leads to the goal , tactical marketing after a good strategic marketing is only the mater of how comfortably you reach the goal.

Good strategic marketing + good Tactical marketing you reach the goal comfortably otherwise its game of survival.

However bad strategic marketing can never lead to success , bad strategic marketing with a good tactical plan is like running efficiently and fast on a wrong mountain.

Therefore Good tactical marketing can never compensate for bad strategic marketing. Strategic marketing is the foundation of organisational success.

Market Segmentation

Market segmentation means cutting markets into slices. Ideally, consumers belonging to the same market segments – or sets of buyers – are very similar to one another with respect to the consumer characteristics deemed critical by management

The segmentation criterion can be one single consumer characteristic, such as age, gender, country of origin, or stage in the family life cycle. Alternatively, it can contain a larger set of consumer characteristics, such as a number of benefits sought when purchasing a product, a number of activities undertaken when on vacation, values held with respect to the environment, or an expenditure pattern.

Different Approaches

1. Concentrated Market Strategy

   Here the organisations who are resource-poor, but are facing fierce competition in the market , can concentrate fully on satisfying the needs of one market segment.

2. Differentiated market strategy

   Here the organisation has enough resources and caters the individual needs of each segment through launching the customised product for each segment.

3. Undifferentiated market strategy

   Here the same product is marketed using the same marketing mix to the entire market. Eg : petrol , white bread. Such an approach may be viable for resource rich organisations or in cases where a new product is introduced and consumers are not yet able to discriminate between alternative products.

4. Hyper segmentation / finer segmentation

   Taking market segmentation to the extreme would mean to actually be able to offer a customised product or service to very small groups of consumers. This approach is referred to as micro marketing or hyper-segmentation. One step further leads to finer segmentation where each consumer represents their own market segment. Finer segmentation approaches are becoming more viable with the rise of eCommerce and the use of sophisticated consumer databases enabling providers of products and services to learn from a person's purchase history about what to offer them next

Market Segmentation Analysis

The process of grouping consumers into naturally existing or artificially created segments such that the consumers who belong to one segment share similar preferences or characteristics regarding a product or service offered.

High level view steps

- Data collection
- Statistical segment extraction
- Segment profiling and describing
- Selecting the target segments and Converting it into strategic marketing decisions and tactical marketing action.

Approaches to Market Segmentation Analysis

1. **Based on Organisational Constraints**
2. **Based on Choice of the Segmentation Variables**

1. Based on Organisational Constraints

Three approaches to market segmentation:

- The quantitative survey-based approach (Segment Revolution)

- The creation of segments from existing consumer classifications (Segment Evolution)
- The emergence of segments from qualitative research (Mutation)

Segment Revolution : We refer to the approach requiring the most radical change in the organisation as *segment revolution*. It is like jumping on an existing sandcastle (or existing market strategy of the organisation in this context) and building a new one. It starts from zero.

Segment Evolution : A less radical approach is that of segment evolution, which is like refining an existing sandcastle (or refining the existing marketing strategy).creating segments from currently targeted sectors and segments. This approach – representing segment evolution rather than revolution – is one of refining and sharpening segment focus

Mutations : The third approach is that of exploratory research pointing to segments. Under this approach, market segments are stumbled upon as part of an exploratory research process possibly being undertaken for a very different purpose initially

2. Based on Choice of the Segmentation Variables

A more technical way of systematising segmentation approaches is to use as a basis the nature of consumer characteristics used to extract market segments.

Something a single piece of information about a customer(one segmentation variable) is used and other times multiple pieces of information (multiple segmentation variables) are used.

Variables such as : age , gender , location , prior purchase , Motives etc

When one single segmentation variable is used, the segmentation approach is referred to as a **commonsense market segmentation**.describes this approach to market segmentation as one that is created without the benefit of primary market research. Managerial intuition, analysis of secondary data sources, analysis of internal consumer databases, and previously existing segments are used to group consumers into different segments.

The term commonsense segmentation implies that users apply their common sense to choose their target segment.

The proactive approach, which exploits multiple segmentation variables. These terms indicate that the nature of the resulting market segments is not known until after the data analysis has been conducted. An alternative term used is that of **data-driven segmentation**. This term implies that the segmentation solution is determined through data analysis, that data analysis creates the solution.

Commonsense and data-driven segmentation are two extremes, Commonsense and data-driven segmentation are two extremes.Rather, various combinations of those approaches are used either sequentially or simultaneously.

Eg: Commonsense/commonsense segmentation results from splitting consumers up into groups using one segmentation variable first. Then, one of the resulting segments is selected and split up further using a second segmentation variable

Data Structures and Data-Driven Market Segmentation

When conducting data-driven market segmentation, data analysts and users of market segmentation solutions often assume that market segments naturally exist in the data. Such naturally occurring segments, it is assumed, need to merely be revealed and described. In real consumer data, naturally existing, distinct and well- separated market segments rarely exist.

Three possible conceptual approaches to data-driven market segmentation:

1. Natural
2. Reproducible
3. Constructive segmentation

Term natural segmentation reflects the traditional view of the given data which assumes that there exists natural segments in the data and the aim of segmentation analysis is to find them.

Term reproducible segmentation refers to the case where natural segments do not exist in the data. But the data is not entirely unstructured either. There exist some pattern or some structure other than the cluster from which useful information can be extracted and segmentation task be carried out.

Term Constructive Segmentation refers to artificially creating segments on the random data where neither cluster structure no any other data structure exists.

Facts are that natural segmentation is extremely rare and nearly 75% of the data sets contain some structure other than cluster structure - which can be exploited to extract market segments. And the entire lack of data structure occurs in only 22% of cases.

Step 1 : Deciding (not) to Segment

Segmenting a market is not free. There are costs of performing the research, fielding surveys, and focus groups, designing multiple packages, and designing multiple advertisements and communication messages . its better not to segment unless the expected increase in sales is sufficient to justify implementing a segmentation strategy.

There are also various implementation barriers to consider.

1. Lack of Senior Management involvement

   The first group of barriers relates to senior management. Lack of leadership, pro-active championing, commitment and involvement in the market segmentation process by senior leadership undermines the success of market segmentation.

2. Lack of resources

   Senior management can also prevent market segmentation to be successfully implemented by not making enough resources available, either for the initial market segmentation analysis itself, or for the long-term implementation of a market segmentation strategy.

3. Organizational barriers

   For successful market segmentation and implementation the following are necessary features the organization must have or should incorporate.

   A. organization’s culture is market-oriented

   B. organization is genuinely willing to change and open to new ideas

   C. organization takes a long-term perspective

   D. communication across organisational units is good

   E. organisation is in the position to make significant (structural) changes

   F. organisation has sufficient financial resources to support a market segmentation strategy

4. Lack of formal marketing function or a qualified marketing expert in the organisation.

Step 2: Specifying the Ideal Target Segment

In Step 2 the organization must determine two sets of segment evaluation criteria.

1. Knock-out criteria. :
    - ➔ Essential, non-negotiable features of segments.
    - ➔ It is a shorter set .
2. Attractiveness criteria. :
    - ➔ To evaluate the relative attractiveness of the remaining market segments.
    - ➔ It is a much longer and much more diverse set.
    - ➔ In a rough way, it is a shopping list for the segmentation team.Members of the segmentation team need to select which of these criteria they want to use to determine how attractive potential target segments are.

1. ## Knock-Out Criteria
    - ➔ To determine if market segments resulting from the market segmentation analysis qualify to be assessed using segment attractiveness criteria.
    - ➔ The segment must be
        a. Homogeneous: members of the segment must be similar to one another.
        b. Distinct: members of the segment must be distinctly different from members of other segments.
        c. large enough: the segment must contain enough consumers to make it worthwhile to spend extra money on customizing the marketing mix for them.
        d. Matching: The organization must have the capability to satisfy segment members' needs.
        e. Reachable: There has to be a way to get in touch with members of the segment in order to make the customized marketing mix accessible to them.
2. ## Attractiveness Criteria

    Apart from Knock-out criteria, attractiveness criteria are available to the segmentation team to consider when deciding which attractiveness criteria are most useful to their specific situation.
    - ➔ It is not binary in nature.
    - ➔ Attractiveness of market segment is rated based on specific criterion which determines whether a market segment is selected as a target segment.
3. ## Implementing a Structured Process

    The most popular structured approach for evaluating market segments in view of selecting them as target markets is the use of **a segment evaluation plot.**

- Segmentation evaluation Plot
    - ★ attractiveness along one axis
    - ★ organizational competitiveness on the other axis

  The segment attractiveness and organizational competitiveness values are determined by the segmentation team.

  The segment evaluation plot cannot be completed in Step 2 of the market segmentation analysis because – at this point – no segments are available to assess yet. But there is a huge benefit in selecting the attractiveness criteria for market segments at this early stage in the process: knowing precisely what it is about market segments that matters to the organization ensures that all of this information is captured when collecting data (Step 3). It also makes the task of selecting a target segment in Step 8 much easier because the groundwork is laid before the actual segments are on the table.

At the end of this step, the market segmentation team should have a list of approximately six segment attractiveness criteria. Each of these criteria should have a weight attached to it to indicate how important it is to the organization compared to the other criteria.

Step 3: Collecting Data

- ### Segmentation Variables

  segmentation variable to refer to the variable in the empirical data used in commonsense segmentation to split the sample into market segments.

1. In commonsense segmentation, the segmentation variable is typically one single characteristic of the consumers in the sample.
2. Market segments are created by simply splitting the sample using this segmentation variable into a segment of women and a segment of men.
3. Describing segments is critical to being able to develop an effective marketing mix targeting the segment. Typical descriptor variables include socio-demographics, but also information about media behaviour, allowing marketers to reach their target segment with communication messages.
4. The difference between commonsense and data-driven market segmentation is that data-driven market segmentation is based not on one, but on multiple segmentation variables. These segmentation variables serve as the starting point for identifying naturally existing, or artificially creating market segments useful to the organisation.
5. In the data-driven case we may, for example, want to extract market segments of tourists who do not necessarily have gender in common, but rather share a common set of benefits they seek when going on vacation.

- Segmentation Criteria

  The term segmentation criterion is used here in a broader sense than the term segmentation variable. The term segmentation variable refers to one measured value, for example, one item in a survey, or one observed expenditure category. The term segmentation criterion relates to the nature of the information used for market segmentation. It can also relate to one specific construct, such as benefits sought.

  The following differences between consumers are the most relevant in terms of market segmentation:

  1. profitability, bargaining
  2. power, preferences for benefits or products, barriers to choice and consumer interaction effects.

- Geographic Segmentation

  Geographic information is seen as the original segmentation criterion used for the purpose of market segmentation. Typically – when geographic segmentation is used – the consumer's location of residence serves as the only criterion to form market segments.

  The key advantage of geographic segmentation is that each consumer can easily be assigned to a geographic unit. As a consequence, it is easy to target communication messages, and select communication channels (such as local newspapers, local radio and TV stations) to reach the selected geographic segments.

  The key disadvantage is that living in the same country or area does not necessarily mean that people share other characteristics relevant to marketers, such as benefits they seek when purchasing a product.

- Socio-Demographic Segmentation

  Socio-demographic segmentation criteria, which include age, gender, income, and education. While socio-demographic segmentation can be useful in some industries, such as luxury goods, cosmetics, baby products, retirement villages, and tourism resort products, it has limitations. Like geographic segmentation, socio-demographic segmentation has the advantage of easily determining segment membership for every consumer. However, in many instances, the socio-demographic criterion is not the cause for product preferences, thus not providing sufficient market insight for optimal segmentation decisions.

- Psychographic Segmentation

  Psychographic segmentation is when people are grouped according to psychological criteria, such as their beliefs, interests, preferences, aspirations, or benefits sought when purchasing a product. Benefit segmentation is the most popular approach, while lifestyle segmentation is based on people's activities, opinions and interests. Psychographic criteria are more complex than geographic or socio-demographic criteria, and most psychographic segmentation studies use a number of segmentation variables. The advantage of the psychographic approach is that it is more reflective of the underlying reasons for differences in consumer behavior. However, the power of the approach depends heavily on the reliability and validity of the empirical measures used to capture the psychographic dimensions of interest.

- Behavioral Segmentation

  The approach of using behavioral similarities or reported behavior to extract segments is an alternative to traditional geographic or demographic variables. This approach can include various behaviors such as prior experience, frequency of purchase, and information search behavior. Using actual behavior rather than stated or intended behavior as the basis of segment extraction is advantageous as it groups people by the similarity that matters most. Behavioural data also avoids the need for the development of valid measures for psychological constructs. However, behavioural data is not always readily available, especially if the aim is to include potential customers who have not previously purchased the product. Examples of behavioural segmentation analyses include using actual expenses of consumers, actual purchase data across product categories, and brand choice behaviour over time.

- Data from Survey Studies

  Survey data is cheap and easy to collect, making it a feasible approach for any organization. But survey data – as opposed to data obtained from observing actual behavior – can be contaminated by a wide range of biases.

  Survey data can be contaminated by biases, which can affect the quality of solutions derived from market segmentation analysis.

  1. Choice of Variables

     Carefully selecting the variables that are included as segmentation variables in data-driven segmentation is critical to the quality of the market segmentation solution. Unnecessary variables can make questionnaires long and tedious, increase the dimensionality of the segmentation problem, and divert the attention of the segment extraction algorithm away from information critical to the extraction of optimal market segments.
  2. Response Options

     Options allowing respondents to answer in only one of two ways, generate binary or dichotomous data. Such responses can be represented in a data set by 0s and 1s.The distance between 0 and 1 is clearly defined and, as such, poses no difficulties for subsequent segmentation analysis. Options allowing respondents to select an answer from a range of unordered categories correspond to nominal variables.
  3. Response Styles

     A wide range of response styles manifest in survey answers, including respondents' tendencies to use extreme answer options (STRONGLY AGREE, STRONGLY DISAGREE), to use the midpoint (NEITHER AGREE NOR DISAGREE), and to agree with all statements. Response styles affect segmentation results because commonly used segment extraction algorithms cannot differentiate between a data entry reflecting the respondent's belief from a data entry reflecting both a respondent's belief and a response style.

4. Sample Size

- Sample size recommendations are not commonly provided for market segmentation analysis.
- A Viennese psychologist recommends a sample size of at least 2p, where p is the number of segmentation variables.
- Qiu and Joe recommend a sample size of at least $10 \cdot p \cdot k$, where k represents the number of segments and the smallest segment should contain a sample of at least $10 \cdot p$ if segments are unequally sized.
- Dolnicar et al. conducted extensive simulation studies with artificial data sets and recommend a sample size of at least $60 \cdot p$ for a typical data scenario and at least $70 \cdot p$ for a more difficult scenario.
- Dolnicar et al. also investigated the effect of market and data characteristics on sample size requirements, finding that larger sample sizes always improve the ability to identify correct market segmentation solutions but the extent to which this is the case varies substantially across market and data characteristics. Correlation between segmentation variables is a particularly challenging characteristic that cannot be compensated for by increasing sample size.

- ## Data from Internal Sources

  Increasingly organizations have access to substantial amounts of internal data that can be harvested for the purpose of market segmentation analysis. Typical examples are scanner data available to grocery stores, booking data available through airline loyalty programs, and online purchase data. The strength of such data lies in the fact that they represent actual behavior of consumers, rather than statements of consumers about their behavior or intentions, known to be affected by imperfect Memory.

  Another advantage is that such data are usually automatically generated and – if organizations are capable of storing data in a format that makes them easy to access– no extra effort is required to collect data.

- ## Data from Experimental Studies

  The response to the advertisement could then be used as a segmentation criterion. Experimental data can also result from choice experiments or conjoint analysis. The aim of such studies is to present consumers with carefully developed stimuli consisting of specific levels of specific product attributes. Consumers then indicate which of the products – characterized by different combinations of attribute levels – they prefer. Conjoint studies and choice experiments result in information about the extent to which each attribute and attribute level affects choice. This information can also be used as a segmentation criterion.

Step 4: Exploring Data

1. First Glimpse at the Data

Data exploration helps to

(1) identify the measurement levels of the variables

(2) investigate the univariate distributions of each of the variables

(3) assess dependency structures between variables

2. Data Cleaning

The first step before commencing data analysis is to clean the data. This includes checking if all values have been recorded correctly.

Reproducibility is important from a documentation point of view and enables other data analysts to replicate the analysis. In addition, it enables the use of the exact same procedure when new data is added on a continuous basis or in regular intervals, as is the case when we monitor

segmentation solutions on an ongoing basis. Cleaning data using code, requires time and discipline, but makes all steps fully documented and reproducible. After cleaning the data set, we save the corresponding data frame using the function save(). We can easily re-load this data frame in future R work sessions using function load().

3. Descriptive Analysis

Descriptive numeric and graphic representations provide insights into the data. Statistical software packages offer a wide variety of tools for descriptive analysis. In R, we obtain a numeric summary of the data with command summary(). This command returns the range, the quartiles, and the mean for numeric variables. For categorical variables, the command returns frequency counts. The command also returns the number of missing values for each variable. Helpful graphical methods for numeric data are histograms, boxplots, and scatter plots. Bar plots of frequency counts are useful for the visualization of categorical variables. Mosaic plots illustrate the association of multiple categorical variables.

Binning: To obtain a histogram, we first need to create categories of values. We call this binning. The bins must cover the entire range of observations and must be adjacent to one another. Usually, they

are of equal length. Once we have created the bins, we plot how many of the observations fall into each bin using one bar for each bin. We plot the bin range on the x-axis and the frequency of observations in each bin on the y-axis.

4.Pre-Processing

4.1.Categorical Variables

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

4.2. Numeric Variables

To balance the influence of segmentation variables on segmentation

results, variables can be standardized. Standardizing variables means transforming them in a way that puts them on a common scale.

4.3 Principal Components Analysis

Principal components analysis (PCA) transforms a multivariate data set containing metric variables to a new data set with variables – referred to as principal components – which are uncorrelated and ordered by importance. The first variable (principle component) contains most of the variability, the second principle component contains the second most variability, and so on.

Step 5 : Extracting Segments

Grouping Consumers

Data-driven market segmentation analysis is exploratory by nature. Consumer data sets are typically not well structured. The result of a market segmentation analysis, therefore, is determined as much by the underlying data as it is by the extraction algorithm chosen.

Different methods for segmentation

➔Distance based methods - Distance-based methods use a particular notion of similarity or distance between observations (consumers), and try to find groups of similar observations.

➔ Model based methods - These methods formulate a concise stochastic model for the market segments.

The size of the available data set indicates if the number of consumers is sufficient for the available number of segmentation variables, the expected number of segments, and the segment sizes. The scale level of the segmentation variables determines the most suitable variant of an extraction algorithm. For distance-based methods, the choice of the distance measure depends on the scale level of the data. The scale level also determines the set of suitable segment-specific models in the model-based approach.

We distinguish directly observable characteristics from those that are only indirectly accessible. Benefits sought are an example of a directly observable characteristic. They are contained directly in the data, placing no restrictions on the segment extraction algorithm to be chosen. An example of an indirect characteristic is consumer price sensitivity.

Distance based methods :

K-Means Clustering is an unsupervised learning algorithm, which groups the unlabelled dataset into different clusters. K defines the number of pre-defined clusters that need to be created in the process, so if K=2, there will be two clusters, and for K=3, there will be three clusters, and so on.

The K Means Algorithm is implemented in the following steps:

1. Decide the number of clusters, i.e. K
2. Select K random points in the dataset. These points will be the centres of each of the K clusters and shall be called Centroids.
3. Assign each data point in the dataset to one of the K centroids, based on the point's distance from each of the centroids.
4. Consider this clustering to be correct and reassign the Centroids to the mean of these clusters.
5. Repeat Step 3. If any of the points change clusters, Go to step 4. Else Go to step 6.
6. Calculate the variance of each of the clusters.

7. Repeat this clustering a specific number of times until the sum of variance of each cluster is minimum.

The Elbow Method

Finding the ideal number of clusters to divide the data into is a critical stage in any unsupervised technique. One of the most prominent techniques for figuring out this ideal value of k is the elbow approach. It is probably the most well-known approach which involves calculating the sum of squares for each cluster size, graphing the results, and identifying the ideal cluster size by looking for an elbow where the slope changes from steep to shallow.

Algorithms with Integrated Variable Selection :

Algorithms focus on extracting segments from data, but sometimes segmentation variables are not carefully selected and contain redundant or noisy variables. To identify suitable segmentation variables, a number of algorithms extract segments while selecting suitable segmen- tation variables. Two such algorithms are biclustering and the variable selection procedure for clustering binary data (VSBD). Factor-cluster analysis is a two-step approach that compresses segmentation variables into factors before segment extraction.

Biclustering Algorithms :

Biclustering is a clustering algorithm used to extract market segments containing consumers who all have a value of 1 for a group of variables. Hartigan (1972) proposed several patterns for direct clustering of a data matrix, but uptake of algorithms such as biclustering, co-clustering, or two-mode clustering was minimal. This changed with the advent of modern genetic and proteomic data, which is characterized by the large numbers of genes. Biclustering algorithms exist for any kind of data, including metric and binary, and differ in how a bicluster is defined. The market segmentation task is to identify tourists who all undertake a subset of all possible activities, and the challenge is to find large groups of tourists who have as many activities in common as possible.

Variable Selection Procedure for Clustering Binary Data (VSBD):

Variable selection procedure for clustering binary data sets, based on the k-means algorithm. The procedure first identifies the best small subset of variables to extract seg-ments, and then

adds additional variables one by one until the increase in within-cluster sum-of-squares reaches a threshold. The number of segments k has to be specified in advance, and the Ratkowsky and Lance index is used to select the number.
The algorithm works by selecting only a subset of observations with size $\varphi \in (0, 1)$ times the size of the original data set. For a given number of variables V, perform an exhaustive search for the set of V variables that leads to the smallest within-cluster sum-of-squares criterion. The value for V needs to be selected small for the exhaustive search to be computationally feasible. To reduce the number of random initialisations, the Hartigan-Wong algorithm is used by default by function kme.

Variable Reduction: Factor-Cluster Analysis :

The concept of factor-cluster analysis, a two-step procedure for data-driven market segmentation analysis. It explains that this approach is only conceptually legitimate in cases where the empirical data results from a validated psychological test battery. However, it is often used because the original number of segmentation variables is too high, which leads to a loss of information and a modified version of the consumer data. The author argues that this approach lacks conceptual legitimisation and comes at a substantial cost, and recommends extracting segments from the space of the original consumer data rather than the space of the resulting factors.Sheppard (1996) suggests that cluster analysis on raw item scores may produce more accurate or detailed segmentation than factor scores. However, empirical evidence suggests that factor-cluster analysis does not outperform cluster analysis using raw data. Dolnicar and Grün (2008) found that factor-analytic models did not outperform clustering of raw data in terms of identifying the correct market segment structure.

Data Structure Analysis :

The process of extracting market segments is exploratory and cannot be validated in the traditional sense of targeting a clear optimality criterion. Instead, validation in market segmentation involves assessing the reliability or stability of solutions across repeated calculations after modifying the data or the algorithm. This approach is referred to as stability-based data structure analysis. Data structure analysis provides insights into the properties of the data, indicating whether natural, distinct, and well-separated market segments exist or not. If they do, they can be easily revealed, but if they do not, exploring a large number of alternative solutions may be necessary to identify the most useful segment(s) for the organization.

Cluster Indices :

Cluster indices are a common approach to provide guidance for data analysts in making critical decisions, such as selecting the number of market segments to extract. There are two types of cluster indices: internal and external. Internal cluster indices are calculated based on a single market segmentation solution, providing insight into particular aspects of the solution, such as the similarity between members within the same segment. External cluster indices, on the other hand, require another segmentation as additional input to measure the similarity between two segmentation solutions. These measures of similarity, such as the Jaccard index, the Rand index, and the adjusted Rand index, are used to compare the stability of market segments extracted from repeated calculations.

Internal Cluster Indices :

Internal cluster indices measure the compactness of clusters by calculating the sum of distances between each segment member and their seg-ment representative.

The k-means algorithm uses an internal cluster index to select the number of market segments for clustering. The scree plot visualizes the sum of within-cluster distances Wk for segmentation solutions containing different numbers of segments k. Ideally, an elbow appears in the scree plot if there is a point (number of segments) in the plot where the differences in sum of the within-clusters distances Wk show large decreases before this point and only small decreases after this point. In consumer data, elbows are not so easy to find in scree plots, and the Ball-Hall index Wk/k is proposed to correct for the monotonous decrease of the internal cluster index with increasing numbers of market segments.

Gorge Plots :

A simple method to assess how well segments are separated is to look at the distance between consumer i and segment representative (centroid, cluster centre) h. This distance can be interpreted as the similarity of consumer i to the representative of segment h. For partitioning methods, segment representatives and distances between con-sumers and segment representatives are directly available. For model-based meth-ods, we use the probability of a consumer i being in segment h given the consumerdata, and the fitted mixture model to assess similarities. Similarity values can be visualised using gorge plots, silhouette plots, or shadow plots.

The plots in the middle column shows the gorge plots for the three-segment solutions extracted using k-means partitioning clustering for these data sets. The x-axis plots similarity values, while the y-axis plots the frequency with which each similarity value occurs. High similarity values indicate that a consumer is very close to the centroid (the segment representative) of the market segment, while low similarities indicate that the consumer is far away from the centroid.

Global Stability Analysis :

Resampling methods are an alternative approach to data structure analysis that can be used for both distance-and model-based segment extraction techniques. They offer insight into the stability of a market segmentation solution across repeated calculations. To assess the global stability of any given seg-mentation solution, several new data sets are generated and a number of segmentation solutions are extracted. To understand the value of resampling methods for market segmentation analysis, it is important to accept that consumer data rarely contain distinct, well-separated market segments like those in the artificial mobile phone data set. Resampling methods provide critical insight into the structure of the data, and it is helpful to develop a systematics of the data before using them.

Consumer data can fall into one of three categories: rarely, naturally existing, distinct, and well-separated market segments exist. If natural segments exist, they can be used for long-term strategic planning and customised marketing mix. If data is entirely unstructured, managerially useful market segments have to be constructed. If segmentation is constructive, the role of the data analyst is to offer potentially interesting segmentation solutions to the user and assist them in determining which of the artificially created segments is most useful to them. There is also a middle option between the worst case and the best case scenario, where the existing structure can be leveraged to extract artificially created segments that re-emerge across repeated calculations. Global stability analysis helps determine which of the concepts applies to any consumer data.

Global stability analysis acknowledges that the sample of consumers and the algorithm used in data-driven segmen-tation introduce randomness into the analysis. Haley (1985) recommends addressing the problem by dividing the sample of respondents into subsamples and extracting market segments independently for each of the subsamples. Dolnicar and Leisch (2010) recommend using bootstrapping techniques to generate new data sets and compute replicate segmentation solutions for different numbers of clusters. Computing the similarity between the resulting solutions for the same number of clusters provides insight into whether natural segments exist in the data.

Segment Level Stability Analysis :

Choosing the globally best segmentation solution does not necessarily mean that it contains the single best market segment. It is important to assess both global and segment level stability of alternative market segmentation solutions to protect against discarding individual segments. Most organisations only need one single target segment.

Segment Level Stability Within Solutions (SLSW) :

Dolnicar and Leisch (2017) propose to assess segmentation solutions based on an approach that determines stability separately for each segment, rather than for the entire market segmentation solution. The criterion of segment level stability within solutions (SLSW) is similar to the concept of global stability, but is computed at segment level, allowing the detection of one highly stable segment in a segmentation solution where several or even all other segments are unstable. Hennig (2007) recommends the following steps: compute a partition of the data, extract k segments, cluster all b bootstrap samples into k segments, assign observations in the original data set to segments Si1,...,Sik for i = 1,...,b, and compute the maximum agreement with the original segments S1,...,Sk as measured by the Jaccard index.
The Jaccard index is the ratio between the number of observations contained in both segments, and the number of observed contained in at least one of the two segments. Segments with higher segment level stability within solutions (SLSW) are more attractive. To demonstrate this, three distinct and well-separated segments are known to exist in the artificial mobile phone data set. Cluster the data into three to eight segments, and label all segments with function relabel. Finally, save the three- and six-cluster solutions into individual objects.
The artificial mobile phone data set from Sect. 7.2.3 is used to demonstrate the procedure of data-driven market segmentation. Three distinct and well-separated segments are known to exist in this data, and if three segments are extracted, the correct segments emerge and segment level stability within solutions (SLSW) is high. However, if more than three segments are clustered, one of the larger natural segments is split up and this split is not stable. To illustrate this, the data first needs to be loaded, then cluster the data into three to eight segments.
Finally, all segments need to be consistently labelled across segmentation solutions using function relabel. Figure 7.41 shows the segmentation solutions for three and six segments, and the global stability of the two segmentation solutions is much more stable than the six-segment solution.

The segment level stability within solutions (SLSW) for the six-segment solution on the right side of Fig. 7.42 indicates that only segment 6 is very stable. This insight is only possible if segment level stability is assessed. For two-dimensional data, data structure is seen by looking at a scatter plot, but typical consumer data is multi-dimensional and cannot be plotted in the same way. Analysing data structure thoroughly when extracting market segments is critically important.

Segment Level Stability Across Solutions (SLSA) :

The second criterion of stability at segment level proposed by Dolnicar and Leisch (2017) is segment level stability across solutions (SLSA). This criterion is used to determine the re-occurrence of a market segment across marketsegmentation solutions containing different numbers of segments. High values of SLSA serve as indicators of market segments occurring naturally in the data, rather than being artificially created. To compare market segmentation solutions, it is necessary to identify which segments in each of the solutions with similar numbers of segments (Pi, Pi+1) are similar to each other and assign consistent labels. An algorithm to renumber P1 such that segments that are similar to segments in P1 get suitable numbers assigned as labels is proposed.
The segment level stability across solutions (SLSA) plot confirms what is seen in the right chart. If more than three segments are extracted from the mobile phone data set, the high-end segment continues to be identified correctly, while the other two (larger) segments gradually get subdivided. The measure of entropy (Shannon 1948) can be used to measure the uncertainty in a distribu-tion. Numerical stability SLSA(Sil) of segment l in these segmentation solution with ki segments is defined bySLSA(Sil) = 1 −pj log pjlog(ki−1) . A value of 0 indicates minimal stability and 1 indicates maximal stability.

Step 6 : Profiling Segments

Identifying Key Characteristics of Market Segments:

The purpose of the profiling stage is to learn more about the market segments that were produced by the extraction process. Profiling is only necessary when using data-driven market segmentation. The segment profiles are established for common sense segmentation.

Age groupings will undoubtedly be the segments that are produced if, for instance, age is utilised as the segmentation variable for the commonsense segmentation. In light of this, Step 6 is not required when using common sense segmentation.
In the case of data-driven segmentation, the scenario is very different: users of the segmentation solution may have chosen to extract segments based on benefits sought by customers. Nevertheless, the distinguishing traits of the resulting market categories are unknown until the data has been analysed. The purpose of profiling is to determine these identifying traits of market segments in relation to the segmentation variables. Profiling entails describing each market segment both on its own and in light of other market segments. When asked about their vacation plans, the majority of winter visitors to Austria say they are going alpine skiing. Alpine skiing might define a market niche, but it might not set that category apart from others.

Traditional Approaches to Profiling Market Segments

To cope with the enormity of this assignment, information regarding the statistical significance of the difference between segments for each of the segmentation factors is supplied at times. This strategy, however, is not statistically correct. Segment membership is determined directly from the segmentation variables, and segments are formed in such a way that they are maximally diverse, preventing the use of normal statistical tests to assess the significance of differences.

Segment Profiling with Visualisations

Visualisations are important in the data-driven market segmentation process for inspecting one or more segments in depth for each segmentation solution. Statistical graphics make it easier to analyse segment profiles. They also make determining the efficacy of a market segmentation option easier. Data segmentation always results in a huge number of potential alternatives. Choosing one of the potential options is an important decision.
The segment profile plot is a type of plot known as a panel. One segment is represented by each of the six panels. The segment profile plot displays the cluster centres for each segment

(centroids, representatives of the segments). The segmentation variables' overall mean values across all observations in the data set are represented by the identical dots in Fig. 8.2's six panels.

A segment separation plot can be used to visualise segment separation. The segment separation plot shows the overlap of segments for all pertinent dimensions of the data space.
When there are few segmentation variables, segment separation plots are extremely straightforward; but, as the number of segmentation factors rises, they become more complicated. Segment separation plots, however, give data analysts and users a fast overview of the data condition and the segmentation solution even in these complex settings.

Parameter that chooses principal components 2 and 3, and axes' labels are assigned by xlab and ylab. By including the projected segmentation variables' directions, the projAxes() function improves the segment separation plot. The improved form combines the strengths of perceptual maps and the segment separation plot.
Owing to market segment overlap (and the n = 1000 sample size), the figure in Fig. 8.5 is disorganised and challenging to interpret. A cleaner version can be achieved by changing the colour scheme (argument col), eliminating the observations (points = FALSE), and highlighting only the inner portion of each segment (hull.args = list(density = 10), where density determines how many lines shade the area).

Step 7: Describing Segments

Developing a Complete Picture of Market Segments

Segment profiling is about understanding differences in segmentation variables across market segments.
step. The only difference
is that the variables being inspected have not been used to extract market segments. Rather, in Step 7 market segments are described using additional information available about segment members. If committing to a target segment is like a marriage, profiling and describing market segments is like going on a number of dates to get to know the potential spouse as well as possible in an attempt to give the marriage the best possible chance, and avoid nasty surprises down the track.

Using Visualizations to Describe Market Segments

Using graphical statistics to describe market segments has two key advantages:

it simplifies the interpretation of results for both the data analyst and the user, and integrates information on the statistical significance of differences, thus avoiding the over-interpretation of insignificant differences.

- Nominal and Ordinal Descriptor Variables

    When describing differences between market segments in one single nominal or ordinal descriptor variable, the basis for all visualisations and statistical tests is a cross-tabulation of segment membership with the descriptor variable.
    Moreover, mosaic plots can display tables with more than two descriptors.variables and incorporate inferential statistical components. This facilitates interpretation. Cellular colours can show regions where recorded frequencies diverge from

    predicted frequencies if the variables are considered to be independent. Based on the standardised difference between the expected and observed frequencies, cell colours are determined. Observations are lower than predicted when there are negative disparities. They have a crimson colour. Positive differences indicate that observed frequencies are higher than predicted. They have a blue colour. The color's saturation reveals the precise value of the standardised difference.

- Metric Descriptor Variables

    In the context of segment description, this R package can display the age distribution of all segments comparatively. Or visualise the distribution of the (original metric) moral obligation scores for members of each segment.
    To have segment names (rather than only segment numbers) displayed in the plot,
    we create a new factor variable by pasting together the word "Segment" and the segment numbers from C6. We then generate a histogram for age for each segment.
    Argument as.table controls whether the panels are included by starting on the top left (TRUE) or bottom left (FALSE, the default).
    We can gain additional insights by using a parallel box-and-whisker plot; it shows the distribution of the variable separately for each segment. We create this parallel
    box-and-whisker plot for age by market segment in R with the following command:
    R> boxplot(Age ~ C6, data = vacmotdesc,
    + xlab = "Segment number", ylab = "Age")

Testing for Segment Differences in Descriptor Variables

Simple statistical tests can be used to formally test for differences in descriptor variables across market segments. The simplest way to test for differences is to run a series of independent tests for each variable of interest. The outcome of the segment extraction step is segment membership, the assignment of each consumer to one market segment. Segment membership can be treated like any other nominal variable.
The association between the nominal segment membership variable and another nominal or ordinal variable (such as gender, level of education, country of origin) is visualized in Sect. 9.2.1 using the cross-tabulation of both variables as basis for the mosaic plot. The appropriate test for independence between columns and rows of a table is the $\chi 2$-test.
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.
The most popular method for testing for significant differences in the means of more than two groups is Analysis of Variance (ANOVA).

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. 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 (x1,...,xp) and variance σ2. The relationship between the dependent variable y and the independent variables x1,...,xp is given by:
$y = \beta_0 + \beta_1 x_1 + ... + \beta_p x_p + ,$
In linear regression models, regression coefficients express how much the dependent variable changes if one independent variable changes while all other independent variables remain constant. The linear regression model assumes that changes caused by changes in one independent variable are independent of the absolute level of all independent variables.
In the linear regression model, the mean value of y given x1,...,xp is modeled by the linear function:

$E[y|x_1,...,x_p] = \mu = \beta_0 + \beta_1 x_1 + ... + \beta_p x_p.$
Generalized 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 modeled with a linear function:
$g(\mu) = \eta = \beta_0 + \beta_1 x_1 + ... + \beta_p x_p.$

- Binary Logistic Regression

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

  $g(\mu) = \eta = \log \mu$
  $1 - \mu$
  The intercept in the linear regression model gives the mean value of the dependent variable if the independent variables x1,...,xp all have a value of 0.

In binomial logistic regression, the intercept gives the value of the linear predictor η if the independent variables x1,...,xp all have a value of 0. The probability of being in segment 3 for a respondent with age 0 and a low moral obligation value is calculated by transforming the intercept with the inverse link function, in this case the inverse logit function:

$g^{-1}(\eta) = \exp(\eta)$
$1 + \exp(\eta)$

- Multinomial Logistic Regression

Multinomial logistic regression can fit a model that predicts each segment simul-taneously. Because segment extraction typically results in more than two market segments, the dependent variable y is not binary. Rather, it is categorical and assumed to follow a multinomial distribution with the logistic function as link function.
With function Anova() we assess if dropping a single variable significantly reduces model fit. Dropping a variable corresponds to setting all regression coefficients of this variable to 0. This means that the regression coefficients in one or several columns of the regression coefficient matrix corresponding to this variable are set to 0. Function Anova() tests if dropping any of the variables significantly reduces model fit.
We assess the predictive performance of the fitted model by comparing the predicted segment membership to the observed segment membership.

- Tree-Based Methods

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

The disadvantage is that results are frequently unstable.
Small changes in the data can lead to completely different trees.
The tree approach uses a stepwise procedure to fit the model. At each step, consumers are split into groups based on one independent variable. The aim of the split is for the resulting groups to be as pure as possible with respect to the

dependent variable. This means that consumers in the resulting groups have similar values for the dependent variable. In the best case, all group members have the same value for a categorical dependent variable. Because of this stepwise splitting procedure, the classification and regression tree approach is also referred to as recursive partitioning.

Step 8: Selecting the Target Segments

The Targeting Decision :

Once a business has segmented the market into distinct groups of customers, the next step is to decide which segments to target. Targeting involves selecting one or more segments to focus on with specific marketing strategies and tactics. Here are some factors that businesses should consider when making targeting decisions based on market segmentation:

1. **Segment size and growth potential:** Businesses should consider the size of each segment and its growth potential when deciding which segments to target. Larger segments with high growth potential may offer greater opportunities for revenue growth and profitability.

2. **Segment profitability:** Businesses should also consider the profitability of each segment when making targeting decisions. Some segments may be more profitable than others, based on factors such as the cost of customer acquisition and the lifetime value of customers in each segment.

3. **Segment accessibility:** Businesses should consider the ease of reaching and engaging with customers in each segment. Some segments may be more accessible than others, based on factors such as geographic location, communication preferences, or purchasing behaviors.

4. **Competitive intensity:** Businesses should also consider the level of competition within each segment when making targeting decisions. Some segments may be highly competitive, with many businesses vying for customers' attention, while other segments may be less competitive and offer greater opportunities for differentiation and market share growth.

Market Segment Evaluation :

Market segment evaluation is a critical process in market segmentation that involves the analysis and assessment of potential target markets for a company's products or services. This process aims to identify the most promising market segments based on their attractiveness, accessibility, and viability for the company.

To evaluate a market segment, companies typically consider several factors, including:

1. **Size of the segment:** The total number of potential customers in the segment and the potential revenue the company can generate from this segment.

2. **Growth potential:** The expected growth rate of the segment over time and the future demand for the company's products or services.

3. **Competition:** The intensity of competition within the segment and the company's ability to compete effectively against existing players.

4. **Profitability:** The potential profit margins and return on investment from serving the segment.

5. **Accessibility:** The ease of reaching and serving the segment, including factors such as geographic location, distribution channels, and communication channels.

6. **Needs and preferences:** The unique needs, preferences, and buying behaviors of customers within the segment, and the company's ability to meet these needs effectively.

By evaluating these factors and other relevant criteria, companies can identify the most attractive and viable market segments to target, develop effective marketing strategies, and tailor their products or services to meet the specific needs and preferences of each segment. This can help companies increase their market share, profitability, and customer loyalty over time.

Step 9: Customizing the Marketing Mix :

Implications for Marketing Mix Decisions

Market segment evaluation has significant implications for marketing mix decisions, as it helps companies develop a more targeted and effective marketing strategy for each market segment. Here are some of the key implications:

1. Product:
   Develop different product variations or features to meet the unique needs and preferences of each segment.
   Use different packaging or branding to appeal to different customer segments.
   Position the product differently to resonate with each segment.

2. Price:
   Determine the appropriate pricing strategy for each segment based on customer willingness to pay, price sensitivity, and competition.

Use different pricing models or offer discounts or promotions to specific segments.
Adjust pricing based on the perceived value of the product or service to each segment.

3. Place:
   Identify the most effective distribution channels and locations for reaching each segment.
   Use different retail channels, online platforms, or physical locations based on the accessibility and preferences of each segment.
   Tailor the distribution strategy to meet the unique needs of each segment.

4. Promotion:
   Develop targeted marketing messages and promotions that resonate with each segment.
   Use different communication channels, messages, or creative elements to appeal to different customer segments.
   Adjust the promotional mix based on the preferences and behavior patterns of each segment.

Overall, market segment evaluation provides companies with a better understanding of their target market segments and helps them make more informed marketing mix decisions. By tailoring their product, price, place, and promotion strategies to the specific needs and preferences of each segment, companies can increase customer satisfaction, loyalty, and profitability over time.

Market Segmentation Case Study on McDonalds Dataset

Kindly refer to any of the following GitHub links for the complete code implementation.

Name	Github Link
Nisarg Patel	Source File
Gaurav Deore	Source File
Saniya Shekh	Source File
Harsh Priyam	Source File
Aradhya Kanth	Source File

Conclusion

It was noted in this study that market segmentation is regarded essential by marketing practitioners for various reasons, including targeting, proposition development, price formulation and developing of mass communication. Though being conceptualised as simple in its rationale, the process of segmentation is not necessarily easy and it requires various considerations that should be taken into account. From the literature it is evident that many marketers are expressing concern about implementation and the integration of segmentation into marketing strategy. To address this, priorities in the area of future segmentation research include the selection and incorporation of new variables into segmentation models, as well as developing new and innovative segmentation strategies.

Using market segmentation, companies are able to identify their target audiences and personalize marketing campaigns more effectively. This is why market segmentation is key to staying competitive. It allows you to understand your customers, anticipate their needs, and seize growth opportunities. This powerful technique allows you to improve your decision-making, marketing efforts, and improve your company’s bottom line.

The key to successful market segmentation remains data quality; therefore, you need to pick your data provider after doing your due diligence, ensuring that you have access to the latest industry information in accessible and easy-to-understand formats.

References

[1] https://rb.gy/1azcm9

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

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