**Summarized notes**

*Submitted by: Srivathsa SH*

*On: 26-10- 2024*

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

Market segmentation is the process marketing managers use to identify and select target markets for their products by dividing a broad market into smaller, more specific groups, or segments, based on shared characteristics. Smith (1956) was the first to suggest using segmentation as a marketing strategy, defining it as breaking down a diverse market into smaller, similar groups. This helps businesses understand their customers better.

***1.1 Implications of Committing to Market Segmentation***

While market segmentation is a valuable strategy for organizations, it’s not universally the best choice. Companies must consider the long-term implications before committing significant time and resources, as effective segmentation. It requires substantial organizational changes and investments in research, surveys, and tailored products and marketing messages; these costs must be justified by expected increases in sales (Cahill, 2006). Organizations may need to develop new products, adjust pricing, modify distribution channels, and enhance communication strategies, which could require changes in their internal structure to focus on different market needs. This decision should be made at the highest levels of the organization and communicated clearly throughout the company to ensure everyone is clear with the terms.

***1.2 Implementation Barriers***

1. **Senior Management Barriers**: Lack of leadership and commitment from senior management can undermine segmentation efforts.
2. **Need for Recognition**: Senior leaders must recognize the need for segmentation, understand the process, and show active interest (McDonald and Dunbar, 1995).
3. **Resource Allocation**: Insufficient resources allocated for both initial analysis and long-term implementation can hinder success.
4. **Organizational Culture Barriers**: A lack of market or consumer orientation within the organization can prevent effective segmentation.
5. **Resistance to Change**: Resistance to change, poor communication, and office politics can obstruct new ideas and creative thinking (Dibb and Simkin, 2008).
6. **Lack of Training**: If senior management and the segmentation team lack understanding of market segmentation principles, the effort is likely to fail.
7. **Absence of a Formal Marketing Function**: Organizations without a qualified marketing expert or formal marketing structure may struggle with implementation.
8. **Objective Restrictions**: Limited financial resources can restrict the ability to pursue segmentation effectively.
9. **Proactive Barrier Identification**: Many barriers can be identified early in the segmentation study and addressed proactively; if barriers cannot be removed, organizations should consider abandoning the segmentation effort.

***1.3 Step 1 Checklist***

This first checklist includes not only tasks, but also a series of questions which, if not answered in the affirmative, serve as knock-out criteria.

|  |  |
| --- | --- |
| Is the organization’s culture market-oriented? | If yes, proceed; if no, reconsider. |
| Is the organization genuinely willing to change? | If yes, proceed; if no, reconsider. |
| Does the organization take a long-term perspective? | If yes, proceed; if no, reconsider. |
| Is the organization open to new ideas? | If yes, proceed; if no, reconsider. |
| Is communication across organizational units effective? | If yes, proceed; if no, reconsider. |
| Is the organization in a position to make significant structural changes? | If yes, proceed; if no, reconsider. |
| Does the organization have sufficient financial resources to support a segmentation strategy? | If yes, proceed; if no, reconsider. |
| Is there visible commitment to market segmentation from senior management? |  |
| Is there active involvement of senior management in the market segmentation analysis? |  |
| Is the market segmentation concept fully understood? | If not, conduct training. |
| Are the implications of pursuing a market segmentation strategy fully understood? | If not, conduct training. |

**Step 2: Specifying the Ideal Target Segment**

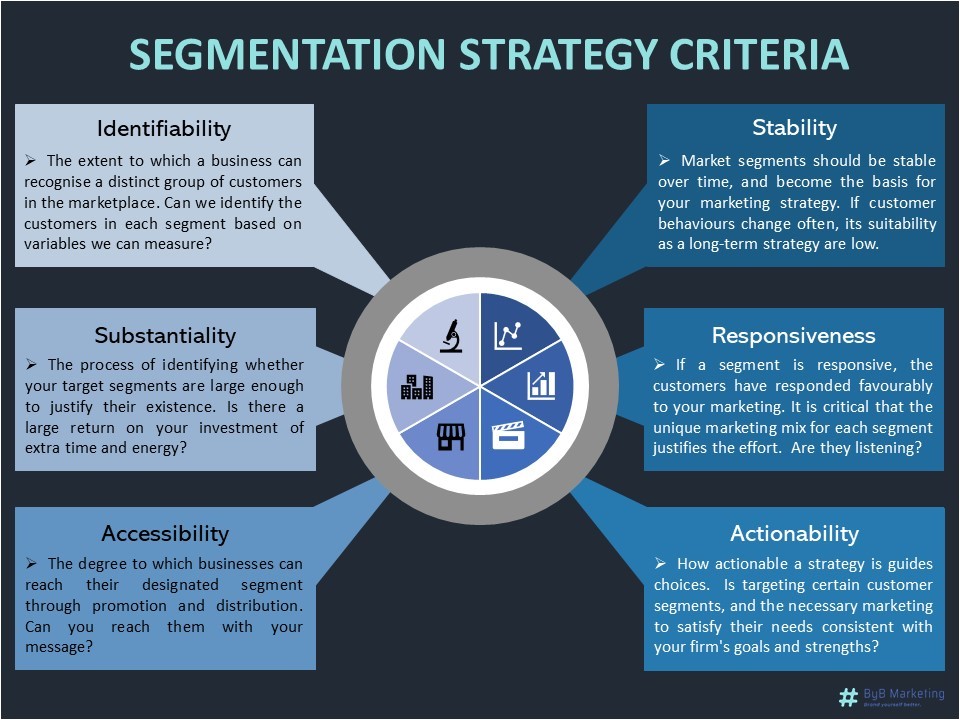
***2.1 Segment Evaluation Criteria***

The third layer of market segmentation analysis relies heavily on user input throughout the process, rather than limiting it to initial briefings or final marketing mix development. User involvement is crucial in most stages, especially after an organization decides to explore a segmentation strategy.

In Step 2 of the analysis, the organization must significantly contribute by establishing two sets of criteria for evaluating segments:

1. **Knock-Out Criteria**: These are essential, non-negotiable attributes that a segment must possess to be considered for targeting. If a segment does not meet these criteria, it is automatically excluded.
2. **Attractiveness Criteria**: Used to assess the appeal of segments that meet the knock-out criteria, these criteria help evaluate and rank the remaining segments for potential targeting.

While literature often discusses a variety of possible evaluation criteria, it does not always clearly differentiate between knock-out and attractiveness criteria. A selection of these criteria is detailed in Table,

Members of the segmentation team need to select which of these criteria they want to use to determine how attractive potential target segments are. The segmentation team also needs to assess the relative importance of each attractiveness criterion to the organisation

***2.2 Knock-Out Criteria***

Knock-out criteria are essential for assessing whether market segments identified in the analysis qualify for further evaluation using attractiveness criteria, the key knock-out criteria include:

* **Homogeneity**: Segment members should be similar.
* **Distinctiveness**: The segment must differ from others.
* **Size**: The segment should be large enough to justify a customized marketing approach.
* **Alignment with Strengths**: The organization must be capable of meeting segment needs.
* **Identifiability**: Segment members should be easily recognizable in the marketplace.
* **Accessibility**: There should be a way to effectively reach segment members.

These criteria must be understood by senior management, the segmentation team, and the advisory committee.

***2.3 Attractiveness Criteria***

In market segmentation analysis, attractiveness criteria are used to evaluate potential segments beyond binary compliance. Each segment is rated on a scale for various criteria rather than simply being categorized as attractive or unattractive. This detailed evaluation helps to provide a clearer view of how appealing each market segment is. In the end, the overall attractiveness ratings for all criteria guide the decision on which segments will be chosen for targeting in the marketing strategy.

***2.4 Implementing a Structured Process***

The literature emphasizes the importance of following a structured approach when assessing market segments. One popular method is to use a segment evaluation plot, which displays segment attractiveness on one axis and organizational competitiveness on the other. The values for attractiveness and competitiveness are determined by the segmentation team, as there is no universal set of criteria applicable to all organizations.

To establish these criteria, various potential factors related to segment attractiveness and organizational competitiveness must be examined, and a consensus reached on the most critical ones. It is advisable to limit the number of key factors to no more than six. This task should ideally be done by a small core team that drafts initial proposals and presents them to a broader advisory committee with representatives from different organizational units. This inclusion is important as various units offer unique perspectives and will be impacted by the segmentation strategy.

While the segment evaluation plot cannot be completed in the early stages of segmentation analysis due to the absence of available segments, selecting attractiveness criteria early is beneficial. It ensures that relevant information is captured during data collection, facilitating target segment selection later in the process.

By the end of the criteria selection step, the team should have a list of around six attractiveness criteria, each assigned a weight to indicate its relative importance. Team members typically allocate a total of 100 points among the criteria, followed by negotiations to reach conclusion. Approval from the advisory committee is recommended to ensure that multiple viewpoints are considered in defining the segment attractiveness criteria.

***2.5 Step 2 Checklist***

| **Task Number** | **Task Description** |
| --- | --- |
| 1 | Convene a segmentation team meeting. |
| 2 | Agree on knock-out criteria for eliminating segments. |
| 3 | Present knock-out criteria to the advisory committee. |
| 4 | Research criteria for assessing segment attractiveness. |
| 5 | Discuss and agree on a maximum of six attractiveness criteria. |
| 6 | Distribute 100 points across the selected criteria. |
| 7 | Finalize weightings for the criteria collaboratively. |
| 8 | Present selected criteria and weights to the advisory committee. |

**Step 3: Collecting Data**

***3.1 Segmentation Variable***

Segmentation variables are key characteristics from empirical data that help create and describe market segments. Commonsense segmentation uses a single variable, such as gender, while data-driven segmentation utilizes multiple variables, like benefits sought, to identify meaningful segments. The quality of this empirical data is vital for accurately assigning individuals to segments and developing effective marketing strategies. Data can be sourced from surveys, observational studies, and experiments, but should reflect true consumer behaviour. Overall, high-quality empirical data is essential for successful market segmentation analysis in both commonsense and data-driven approaches.

***3.2 Segmentation Criteria***

Before extracting market segments, organizations must choose an appropriate segmentation criterion, which encompasses more than just a single variable and requires market knowledge. Common criteria include geographic, socio-demographic, psychographic, and behavioural factors. Bock and Uncles (2002) highlight the importance of considering profitability and consumer preferences. While there are many potential criteria, the simplest approach is often best.

| **Segmentation Type** | **Criteria** | **Advantages** | **Disadvantages** | **Examples** |
| --- | --- | --- | --- | --- |
| **Geographic Segmentation** | Location of residence (e.g., country, city) | Easy to assign consumers to segments; facilitates local targeting | May not capture shared preferences among consumers in the same area | National tourism campaigns (e.g., Austria targeting neighbouring countries) |
| **Socio-Demographic Segmentation** | Age, gender, income, education | Clear segment membership; can explain some product preferences | Provides limited insight (Haley, 1985); demographics explain ~5% of behaviour | Luxury goods marketing targeted at high-income individuals |
| **Psychographic Segmentation** | Beliefs, interests, preferences, aspirations | Reflects deeper motivations for consumer behaviour | More complex to determine segment memberships; reliant on valid measures | Travel companies targeting cultural tourists based on interests |
| **Behavioural Segmentation** | Purchase behaviour (e.g., frequency, spending) | Uses actual behaviour for segmentation; highly relevant insights | Data may be unavailable for potential customers who haven’t purchased yet | Analysing actual consumer spending patterns for targeted promotions |

***3.3 Data from Survey Studies***

Data from survey studies are commonly used for market segmentation analysis due to its low cost and ease of collection. However, survey data can be biased and negatively impact the quality of segmentation solutions. Careful selection of variables is crucial in both commonsense and data-driven segmentation to ensure optimal market segmentation solutions.

***3.3.1 Choice of variable***

* Carefully selecting the variables for segmentation is crucial for both commonsense and data-driven segmentation quality.
* In data-driven segmentation, all relevant variables related to the segmentation criterion must be included, while unnecessary ones should be avoided to prevent respondent fatigue and lower quality responses.
* Noise variables, which do not contribute to identifying correct market segments, can make extracting the correct solution challenging. This can be avoided by asking necessary and unique questions, without including redundant ones.
* Redundant questions, common in survey research, can interfere with segment extraction algorithms.
* Developing a good questionnaire involves both exploratory and qualitative research to ensure all important variables are included. This process helps in understanding people's beliefs and categorizing insights for questionnaire inclusion.

***3.3.2 Response Options***

* Response options in surveys play a crucial role in determining the scale of data available for analysis. Options that allow respondents to answer in only two ways generate binary data, while those that allow selection from a range of unordered categories correspond to nominal variables.
* Metric data, on the other hand, are generated when respondents indicate a number, such as age. Metric data are ideal for segmentation analysis as they allow for various statistical procedures to be applied.
* Ordinal data, which are commonly used in surveys, arise when respondents are asked to express their agreement using a limited number of ordered answer options. While this format captures fine nuances of responses, it may pose challenges in applying standard distance measures.
* Providing binary or metric response options to respondents is preferred to avoid complications in segmentation analysis
* In some cases, binary response options have been shown to outperform ordinal options, especially when formulated in a level-free manner.

***3.3.3 Response Style***

* Survey data can be influenced by biases, including response bias, where respondents consistently answer in a certain way regardless of the question.
* Various response styles, such as always agreeing or using extreme options, can impact segmentation results.
* For instance, a segment of tourists who say yes to all spending items may seem lucrative, but it could simply be a response style rather than actual spending behaviour. To avoid misinterpretation, it is essential to identify and address response styles during data collection, especially in marketing segmentation
* Additional analysis may be needed to verify genuine market segments and exclude biased responses

***3.3.4 Sample size***

Sample sizes are not typically recommended for market segmentation analysis, but it is crucial that the sample size is adequate in order to determine the correct number and nature of segments within the data. Various studies have explored sample size recommendations for different segmentation variables and segments within the data set.

Formann (1984) suggested that the sample size should be at least 2p (or better five times 2p), where p is the number of segmentation variables, specifically for latent class analysis with binary variables. Qiu and Joe (2015) proposed a recommendation of at least ten times the number of segmentation variables times the number of segments in the data for constructing artificial data sets to study clustering algorithms. Dolnicar et al. (2014) conducted simulation studies and found that a sample size of at least 60p is ideal for correctly identifying segments, with no major improvements beyond a sample size of 70p in more complex scenarios.

Dolnicar et al. (2016) expanded this research to account for market characteristics that impact segmentation algorithms, such as the number and size of market segments, segment overlap, and survey data quality issues like response biases and correlation between items. Larger sample sizes consistently improve an algorithm's ability to identify the correct segmentation solution, but the impact varies based on market and data characteristics. For instance, uncorrelated segmentation variables lead to better segment recovery, while high correlation between variables makes the task challenging even with a larger sample size.

The recommendation from Dolnicar et al. (2016) is to ensure the data contains at least 100 respondents for each segmentation variable to achieve accurate segmentation results. It is crucial to collect high-quality unbiased data with the right items, no unnecessary items, no correlated items, high-quality responses, binary or metric data, no response styles, appropriate sample size, and no correlated items for successful market segmentation analysis.

In conclusion, having a sufficient sample size is essential for accurate market segmentation analysis, and various factors such as market characteristics and data quality can impact the effectiveness of segmentation algorithms. Adhering to recommendations for sample size and data quality parameters can help ensure reliable segmentation results.

***3.4 Data from Internal Sources***

Organizations use internal data, such as scanner records from grocery stores, airline booking data, and online purchase histories, for market segmentation. This data shows real consumer behaviour, which is more reliable than people's self-reported habits that can be biased. Since this data is usually collected automatically, it’s easy to access. However, a major drawback is that it often comes from existing customers, which might not accurately represent potential new customers and their preferences.

***3.5 Data from Experimental Studies***

Experimental data comes from controlled studies, either in the field or in the lab. This includes consumer reactions to specific advertisements and results from choice experiments, where consumers choose between different products with various features. This data helps marketers understand which product attributes are most appealing to consumers, allowing them to create more targeted marketing strategies based on consumer preferences.

**Step 4: Exploring Data**

***4.1 A First Glimpse at the Data***

Exploratory Data Analysis (EDA) is a crucial step following data collection that focuses on cleaning and preprocessing the dataset to facilitate effective market segmentation. It aims to identify measurement levels of the variables, investigate their univariate distributions, and assess dependency structures between them. By employing commands to review column names, dimensions, and summaries, researchers can glean valuable insights into the data distribution, ultimately guiding the selection of appropriate segmentation methods based on the characteristics of the dataset.

***4.2 Data Cleaning***

Data cleaning is the initial and essential step in data analysis, focused on ensuring the accuracy and consistency of collected data. This process involves verifying that all values are correctly recorded and that categorical variables utilize consistent labels. For metric variables, known plausible ranges assist in identifying errors; for instance, age should typically fall between 0 and 110 years, allowing for easy detection of any outliers or erroneous entries. Additionally, for categorical variables like gender, it is crucial to check that only permissible values (e.g., female and male) are present in the dataset, ensuring compliance with the survey design. Any discrepancies discovered during this review should be addressed to maintain data integrity before further analysis. Reordering variables like Income is a key part of data cleaning that ensures reproducibility in data analysis. After cleaning the dataset, it can be saved using the save( ) function and reloaded in future sessions with load( ), enhancing efficiency and data integrity.

***4.3 Descriptive Analysis***

Descriptive analysis is crucial for understanding data, preventing misinterpretation of complex analyses. It utilizes numeric summaries and graphical representations to provide insights into datasets. In R, the summary() function delivers a concise overview, including ranges, quartiles, means for numeric variables, and frequency counts for categorical variables, along with the count of missing values.

Graphical methods such as histograms, boxplots, and bar plots enhance data visualization. Histograms display the distribution of numeric variables, allowing for identification of patterns like unimodality or skewness. Binning is essential for creating histograms, where equal-length bins cover the observation range. The lattice package in R can be used for finer bins that provide more detailed insights, such as identifying bimodal distributions.

Boxplots summarize distributions effectively by showcasing the five-number summary: minimum, first quartile, median, third quartile, and maximum. They can indicate data skewness; for instance, a right-skewed distribution shows a median close to the first quartile. Outliers, such as an unusually high age, can affect boxplot representations. R typically limits whisker lengths to prevent skewed data representation, defining outliers as points beyond 1.5 times the interquartile range, thus ensuring that valuable information about outliers is preserved in the analysis.

***4.4 Pre-Processing***

***4.4.1 Categorical Data Pre-Processing***

Pre-processing of categorical data often involves two main procedures: merging levels and converting to numeric formats.

1. **Merging Levels**: This technique is applied when a categorical variable has too many distinct categories, making analysis challenging. By consolidating similar categories, data analysis becomes more straightforward.
2. **Converting to Numeric Variables**: Categorical variables, especially ordinal data, can often be converted to numeric values if it is reasonable to assume that the distances between categories are equal. Consequently, it may be safer to opt for binary options (e.g., Yes/No), which are less susceptible to interpretation issues and don't require extensive pre-processing.

***4.4.2 Numeric Data Pre-Processing***

Numeric data may require standardization to ensure that various variables have equal influence in distance-based analyses. Standardization typically involves subtracting the mean and dividing by the standard deviation, resulting in normalized values with a mean of 0 and a standard deviation of 1. This process can be executed in R using the scale() function. When datasets contain outliers—observations significantly different from others—robust methods, such as using the median and interquartile range, are recommended for standardization to avoid skewed results and ensure more reliable analyses.

***4.5 Principal Components Analysis***

Principal Components Analysis (PCA) is a statistical technique used to transform a multivariate dataset with metric variables into a new set of variables called principal components, which are uncorrelated and ordered by their importance in explaining variance. The first principal component captures the most variability, followed by the second, and so forth. Although PCA generates the same number of new variables as there were original ones, it enables a different perspective on the data without changing its dimensionality.

PCA operates on the covariance or correlation matrix of numeric variables. When all variables are measured on the same scale, the choice between these matrices is not critical. However, if there are differing data ranges, the correlation matrix should be used, effectively standardizing the data.

The primary application of PCA is to reduce the dimensionality of high-dimensional data for visualization purposes, often focusing on the first two or three principal components that reflect the most variation. The output of PCA includes the standard deviations of the principal components and the proportion of variance they explain. For instance, the first principal component may explain 18% of the variance, while the second explains 9%, together accounting for 27% of the total variation.

In practical applications, PCA can reveal relationships among variables. For example, in an analysis of travel motives, some principal components may not effectively differentiate

between certain categories, while others will reflect distinct patterns. This insight is valuable for understanding consumer preferences and motivations.

While using a subset of principal components as segmentation variables is not recommended because it replaces original variables with potentially less informative components, PCA is useful for exploring data and identifying highly correlated variables. These highly correlated variables will have high loadings on the same principal components, signifying redundancy in the information. Insights gained from this exploratory analysis can inform the removal of redundant original variables from the segmentation base. This method still achieves a reduction in dimensionality but retains the original variables, thereby maintaining the richness of the data.

Overall, PCA is a powerful tool for simplifying complex datasets, enhancing visualization, and aiding in exploratory data analysis. However, careful consideration must be given to the dimensionality reduction process to preserve the integrity and utility of the original data.

***4.6 Step 4 Checklist***

The Step 4 checklist involves several key tasks to prepare data for analysis. First, explore the dataset to identify any inconsistencies or systematic contaminations, and clean the data as necessary. Following this, apply appropriate pre-processing steps to ensure the data is ready. It's crucial to assess whether there are enough consumers (at least 100) for each segmentation variable; if the number of variables is excessive, use available methods to select a manageable subset. Additionally, check for correlations among segmentation variables, opting for a subset of uncorrelated variables if needed. Finally, transition the cleaned and pre-processed data to Step 5, where segmentation will be extracted. This comprehensive approach ensures the dataset is well-prepared for further analysis.

Code: https://github.com/srivathsa26/mc-donalds-market-segmentation.git