# **MARKET SEGMENTATION ANALYSIS**

Rajarajeswari T

02-08-2024

**STEP 1:**

**Implications of Committing to Market Segmentation**

Market segmentation is a strategic marketing approach that requires a long-term commitment from organizations. It involves significant changes and investments, such as developing new products, modifying existing ones, and altering pricing and distribution channels. These changes can impact the organization's internal structure, requiring it to organize around market segments rather than products.

The main implication of deciding to use market segmentation is the need for significant and continuous investments. These investments involve expenses for research, surveys, focus groups, various package designs, and different advertisements. According to Cahill (2006), it is important to ensure that the anticipated increase in sales justifies these costs. The choice to adopt market segmentation should be made at the highest executive level and consistently communicated across all levels of the organization.

**Implementation Barriers**

Several obstacles may hinder the successful execution of a market segmentation strategy, including senior management, organizational culture, and operational challenges.

**Senior Management:**

* Lack of leadership and involvement from senior management can undermine the success of market segmentation.
* Insufficient resources allocated by senior management can hinder both the initial analysis and the long-term implementation of the strategy.

**Organizational Culture:**

* Resistance to change, lack of market orientation, poor communication, and office politics.
* Lack of training and understanding of market segmentation among senior management and the tasked team.
* The absence of a formal marketing function or qualified marketing experts, especially in diverse and large organizations, can be a significant obstacle.

**Operational Challenges:**

* Limited financial resources and the inability to make necessary structural changes can restrict market segmentation efforts.
* Process-related issues such as unclear objectives, poor planning, lack of structured processes, and time pressure can hinder the segmentation process.
* Management's reluctance to use complex techniques can be countered by presenting segmentation analysis in an easy-to-understand and visual format.

**STEP 2:**

**Segment Evaluation Criteria**

The third layer of market segmentation analysis relies heavily on user input, which is crucial throughout most stages of the process. In this step, the organization needs to establish two sets of segment evaluation criteria: knock-out criteria and attractiveness criteria. Knock-out criteria are non-negotiable and essential for segment selection, while attractiveness criteria evaluate the relative attractiveness of the segments that meet the knock-out criteria. The segmentation team needs to select and weigh these attractiveness criteria to determine the overall attractiveness of each market segment.

**Knock-Out Criteria**

The knock-out criteria are used to filter out market segments that do not qualify for further evaluation. These criteria, which were initially suggested by Kotler, include substantiality, measurability, accessibility, homogeneity, distinctiveness, size, and organizational capability. It is essential for senior management, the segmentation team, and the advisory committee to understand these criteria. Some criteria, such as size, require further specification, like defining the minimum viable target segment size.

**Attractiveness Criteria**

The attractiveness criteria evaluate the appeal of each market segment based on various factors. These criteria involve rating segments based on their attractiveness concerning each criterion. The segmentation team discusses and chooses the most relevant criteria, which are then utilized to determine the overall attractiveness of each segment in Step 8 of the market segmentation analysis. It's important to develop a list of around six attractiveness criteria, each assigned a weight to indicate its importance.

**Implementing a Structured Process**

A well-defined method for assessing market segments is commonly advised in segmentation literature. The segment evaluation plot is a popular approach, where segment attractiveness is plotted against organizational competitiveness. The criteria for these evaluations need to be discussed and agreed upon by the segmentation team, which should ideally include representatives from various organizational units. This is important for obtaining diverse perspectives and involving stakeholders, as segmentation impacts all units within the organization.

**STEP 3:**

**Segmentation Variables**

Empirical data forms the basis for both common sense and data-driven market segmentation. It is utilized to identify or create market segments and provide detailed descriptions of these segments.

**Segmentation Variables:**

* Commonsense Segmentation:

In common sense segmentation, a single characteristic, known as the segmentation variable, is used to divide the sample into market segments. For instance, using gender as the segmentation variable divides the sample into segments of women and men. Other characteristics, such as age, number of vacations taken, and benefits sought during vacation, act as descriptor variables. These variables are essential for providing detailed descriptions of segments, which is crucial for developing an effective marketing mix.

* Data-Driven Segmentation:

Data-driven segmentation uses multiple variables to identify or create useful market segments. For example, tourists seeking relaxation, culture, and social interactions can form a segment, regardless of their gender. Descriptor variables in this context also include socio-demographic factors and the number of vacations taken.

**Importance of Data Quality:**

High-quality empirical data is essential for accurate segmentation. In common-sense segmentation, data quality ensures correct assignment to segments and accurate segment descriptions, enabling the creation of customized products, pricing, distribution, and communication strategies.

In data-driven segmentation, data quality affects the identification and description of market segments. Good market segmentation analysis requires high-quality empirical data.

**Sources of Empirical Data:**

Empirical data can come from surveys, observations (e.g., scanner data linked to loyalty programs), and experimental studies. Data should ideally reflect actual consumer behavior.

While surveys are common, they can be unreliable, especially for socially desirable behaviors. Thus, various data sources should be considered, with preference given to those reflecting real consumer behavior.

**Segmentation Criteria**

**Segmentation Criteria:** Before beginning to extract market segments and collect data, organizations need to decide on the segmentation criterion. This criterion is broader than a segmentation variable and refers to the type of information used for market segmentation. Common criteria include geographic, socio-demographic, psychographic, and behavioral factors.

**Geographic Segmentation**

* Definition: Divides the market based on consumers' locations (e.g., country, region).
* Advantages: Easy to implement, suitable for targeting specific languages or regions.
* Disadvantages: People in the same area may not share other relevant characteristics; location alone rarely explains product preferences.
* Examples: National tourism organizations and companies like Amazon and IKEA use geographic segmentation to tailor services and products to different regions.

**Socio-Demographic Segmentation**

* Definition: Segments the market based on demographic factors (e.g., age, gender, income).
* Advantages: Easy to determine segment membership; sometimes directly linked to product preferences.
* Disadvantages: Demographics alone often do not explain consumer behavior; can miss deeper insights into preferences and values.
* Examples: Luxury goods (high income), cosmetics (gender), baby products (parents), retirement villages (age).

**Psychographic Segmentation**

* Definition: Groups consumers based on psychological criteria (e.g., beliefs, interests, preferences).
* Advantages: More reflective of underlying reasons for consumer behavior; can provide deeper insights into motivations.
* Disadvantages: Complex to implement; requires reliable and valid measures.
* Examples: Tourism studies often use travel motives as segmentation variables.

**Behavioral Segmentation**

* Definition: Segments based on actual consumer behavior (e.g., purchase frequency, spending habits).
* Advantages: Directly relates to the behavior of interest; avoids the need for developing psychometric measures.
* Disadvantages: Behavioral data may not always be available, especially for potential customers.
* Examples: Analyses based on actual purchase data, brand choice behavior over time.

**Data from Survey Studies**

**Choice of Variables**

* Importance: Critical for the quality of market segmentation.
* Guidelines: Include all relevant variables, avoid unnecessary ones to prevent respondent fatigue and algorithmic challenges.
* Common Issues: Noisy variables and redundant questions can mislead segmentation algorithms.
* Recommendation: Conduct exploratory research to develop a comprehensive yet concise questionnaire.

**Response Options**

* Scales: Use binary or metric response options where possible; avoid ordinal scales due to complications with distance measures.
* Binary Options: Represented by 0s and 1s, suitable for distance measures.
* Metric Data: Allows statistical procedures and is ideal for segmentation.
* Visual Analogue Scales: Recommended for capturing fine nuances without response biases.

**Response Styles**

* Definition: Systematic tendencies in survey responses unrelated to item content (e.g., extreme answers, midpoint usage).
* Impact: Can distort segmentation results; algorithms may misinterpret biased patterns.
* Mitigation: Minimize response styles during data collection; conduct additional analyses to verify segment validity.

**Sample Size**

* Importance: Crucial for accurate market segmentation.
* Guidelines: At least 100 respondents per segmentation variable.
* Challenges: Unequal segment sizes, overlapping segments, and correlated items complicate segment extraction.
* Recommendation: Ensure large, high-quality, unbiased samples to improve algorithm performance.

**Data from Internal Sources**

Internal data encompasses scanner data from grocery stores, booking data from airline loyalty programs, and online purchase data.

Advantages:

- Reflects actual consumer behavior rather than self-reported intentions or behaviors.

- Automatically generated, reducing data collection efforts.

Challenges:

- Potential bias due to over-representation of existing customers.

- Lacks information about potential future customers who may have different consumption patterns.

**Data from Experimental Studies**

Types of Data:

Experimental data can be collected from field experiments, laboratory experiments, choice experiments, or conjoint analyses.

Applications:

- Field/Laboratory Experiments: Used to test responses to advertisements or other stimuli.

- Choice Experiments and Conjoint Analyses: Present consumers with products characterized by various attribute levels to determine preferences and attribute importance.

Advantages:

- Provides insights into consumer preferences and the impact of specific product attributes on choices.

- Can be used as a segmentation criterion to identify distinct market segments based on experimental responses.

**STEP 7:**

**Understanding Market Segments**

Segment profiling involves examining the differences in segmentation variables across market segments to understand them better. This process begins in Step 2 (specifying the ideal target segment) and continues through Step 3 (collecting data) and Step 7 (describing segments). While Step 7 provides a deeper understanding of segments by integrating additional information about segment members, profiling focuses on the segmentation variables used to extract the segments.

**Importance of Segment Description**

Effective segment description is crucial for crafting a targeted marketing mix. For example, if a segment shows a strong interest in nature, understanding their age, income, and vacation behaviors can help tailor marketing strategies. Segment descriptions involve additional variables, such as demographic, psychographic, and behavioral data.

**Using Visualizations for Descriptor Variables**

**Nominal and Ordinal Descriptor Variables**

To visualize differences in nominal and ordinal descriptor variables across segments, the process begins with cross-tabulating segment membership with descriptor variables. For instance, gender distribution across segments can be displayed using stacked bar charts, which show the proportion of males and females in each segment. However, to compare proportions across segments more clearly, mosaic plots are useful as they account for segment size and show proportional differences effectively.

**Metric Descriptor Variables**

For metric descriptor variables, such as income or age, visualizations can reveal associations with segment membership. Examples include:

* Income Distribution: Mosaic plots can show how income levels are distributed across different market segments. For instance, segments motivated by cultural offers may have higher incomes compared to those seeking budget-friendly options.
* Moral Obligation Scores: Analyzing moral obligation to protect the environment can illustrate how different segments value environmental behaviors. For example, nature-loving segments may have higher moral obligation scores compared to segments focused on entertainment.

**Key Insights from Visualizations**

* Gender Distribution: The visual analysis of gender distribution across segments showed no significant differences, suggesting gender balance across market segments.
* Income Levels: There was a noticeable pattern where higher income segments were associated with specific travel motives, such as cultural interests, while lower income segments preferred budget-friendly options.
* Environmental Morality: Segments with strong environmental motivations had higher moral obligation scores, indicating a clear association between environmental values and travel preferences.

**Metric Descriptor Variables and Conditional Plots**

1. **Visualizing Segment Differences:**
   * Lattice and ggplot2 Packages: The R package lattice and ggplot2 offer tools for conditional plots, which divide data into panels for different subsets (e.g., market segments). This helps in comparing metrics like age and moral obligation across segments.
   * Histograms: Histograms created using lattice for age and moral obligation across segments revealed minor differences. Segment-specific histograms are useful but can be challenging for detecting subtle differences.
2. **Box-and-Whisker Plots:**
   * Age Distribution: A box-and-whisker plot for age across segments indicated minimal differences, with segments 5 and 6 showing slightly lower and higher median ages, respectively.
   * Moral Obligation: Parallel box-and-whisker plots for moral obligation showed more variation, with significant differences in moral obligation between segments, particularly highlighting segments 5 and 6 as having higher scores.
3. **Statistical Testing for Differences:**
   * Chi-Squared Test: Used for nominal variables (e.g., gender), with a non-significant result indicating no significant gender distribution differences across segments.
   * ANOVA: Applied to test for differences in means of metric variables (e.g., moral obligation). Results showed significant differences across segments, with segments 5 and 6 differing notably from others.
   * Kruskal-Wallis Test: A non-parametric alternative to ANOVA, used to test for differences in medians across segments.
4. **Pairwise Comparisons:**
   * Pairwise t-Tests: Conducted to identify which segments differ significantly in moral obligation. Results showed that segments 5 and 6 have significantly higher moral obligation scores compared to others.
   * Tukey’s HSD Test: Used for multiple comparisons, indicating that segments 5 and 6 are distinctively higher in moral obligation compared to segments 1 through 4.
5. **Segment Level Stability Across Solutions (SLSA) Plot:**
   * SLSA Plot: This plot, modified to include color-coding based on moral obligation, visually represents stability and moral obligation levels across segments over various solutions. It highlights segments with high moral obligation consistently across different solutions.

**Predicting Segments from Predictor Variable**

The aim is to predict segment membership based on descriptor variables using regression models, with a focus on how well these models identify market segments and which descriptor variables are crucial for segment identification.

**1. Linear Regression Analysis:**

* **Model Structure:** The basic linear regression model assumes a linear relationship between the dependent variable (segment membership) and independent variables (descriptor variables). The model is expressed as:

y=β0+β1x1+…+βpxp+ϵy = \beta\_0 + \beta\_1 x\_1 + \ldots + \beta\_p x\_p + \epsilony=β0​+β1​x1​+…+βp​xp​+ϵ

where ϵ∼N(0,σ2)\epsilon \sim N(0, \sigma^2)ϵ∼N(0,σ2).

* **Application Example:** In R, we fitted a linear regression model to predict age based on segment membership. The output indicated varying mean ages across segments, with the youngest segment having a mean age of 39.4 years and the oldest segment having a mean age of 49.6 years.

**2. Generalised Linear Models (GLMs):**

* **Overview:** GLMs extend linear models to accommodate different distributions for the dependent variable and include a link function to model the mean value of the dependent variable.
* **Binary Logistic Regression:** Used for binary outcomes, where the logit link function maps the probability of success (μ\muμ) to the entire real line. The model is specified as:

g(μ)=log⁡(μ1−μ)=η=β0+β1x1+…+βpxpg(\mu) = \log \left(\frac{\mu}{1 - \mu}\right) = \eta = \beta\_0 + \beta\_1 x\_1 + \ldots + \beta\_p x\_pg(μ)=log(1−μμ​)=η=β0​+β1​x1​+…+βp​xp​

* **Application Example:** We predicted the likelihood of a consumer belonging to segment 3 based on age and moral obligation using binary logistic regression. The results showed that:
  + The probability of segment membership decreased with age.
  + Moral obligation significantly impacted segment membership, with higher levels leading to lower probabilities of being in segment 3.

**3. Model Evaluation and Comparison:**

* **Performance Metrics:** We evaluated model performance using metrics like deviance, AIC, and predicted probabilities. The comparison involved two models:
  + **Basic Logistic Regression Model:** Included age and moral obligation.
  + **Stepwise Selected Model:** Included education, NEP (New Environmental Paradigm), and vacation behavior.
* **Findings:**
  + The stepwise-selected model provided slightly better predictive performance for segment 3 membership.
  + The predicted probabilities for segment 3 were higher in the stepwise model compared to the basic model, though neither model achieved optimal differentiation between segment members and non-members.

**Multinomial Logistic Regression**

1. **Purpose:**
   * Multinomial logistic regression is used for predicting categorical dependent variables with more than two categories (market segments in this case).
   * It fits a model to predict each segment simultaneously, using a logistic function as the link function.
2. **Implementation in R:**
   * The multinom() function from the nnet package is employed for fitting multinomial logistic regression models
3. **Model Output:**

* The model provides regression coefficients for each segment (excluding the baseline category).
* Coefficients reflect the change in log odds for changes in independent variables.
* Model fit can be assessed using the Anova() function, which evaluates if dropping a variable significantly reduces the model fit.

1. **Model Evaluation**:
   * Predictive performance is assessed by comparing predicted segment membership with observed membership.
   * Visualization of predicted probabilities for each segment helps in understanding model performance.
   * Use of plot(allEffects()) to interpret the effect of predictors on segment probabilities.

**Tree-Based Methods**

1. **Purpose**:
   * Classification and Regression Trees (CART) are used for predicting binary or categorical dependent variables.
   * Trees provide variable selection, ease of interpretation, and incorporate interaction effects. They handle large numbers of independent variables but can be unstable with small data changes.
2. **Implementation in R**:
   * The rpart package and partykit package are commonly used for tree construction.
   * For a binary dependent variable, use the ctree() function from partykit:

library("partykit")

tree63 <- ctree(factor(C6 == 3) ~ ., data = vacmotdesc)

* For a categorical dependent variable with more than two categories:

tree6 <- ctree(C6 ~ ., data = vacmotdesc)

1. **Model Output**:

* Trees are visualized showing nodes from splits. Terminal nodes indicate final segment predictions.
* Example code:

plot(tree63)

plot(tree6)

1. **Model Evaluation**:

* The complexity of the tree and the proportion of correct predictions in terminal nodes are evaluated.
* Trees can be adjusted with parameters such as minimum bucket size and criteria for splits.

1. **Examples**:

* For segment 3, a tree was built showing splits based on vacation behavior and obligation.
* For segment 6, a more complex tree showed detailed splits, illustrating higher moral obligation and NEP score.

**STEP 8:**

**1. The Targeting Decision:**

The targeting decision is a critical step where a company selects one or more market segments to focus on. This decision, made after segment profiling and description, is influenced by several factors:

* Knock-out Criteria: Segments that do not meet basic criteria (e.g., size, homogeneity, distinctiveness) are excluded early in the process.
* Segment Attractiveness and Organizational Competitiveness: The remaining segments are assessed based on their attractiveness to the organization and the organization’s competitiveness in addressing the segment’s needs.

The key questions to address are:

1. Which segments are most desirable for the organization to target?
2. How likely is the organization to be chosen by these segments compared to competitors?

**2. Market Segment Evaluation:**

To facilitate decision-making, a decision matrix is used to evaluate and compare market segments. Common matrices include the Boston Matrix, General Electric/McKinsey Matrix, and Directional Policy Matrix.

* Axes of the Matrix:
  + Segment Attractiveness (x-axis): Reflects how desirable the segment is based on criteria such as growth potential, profitability, and fit with organizational goals.
  + Organizational Competitiveness (y-axis): Measures how well the organization can meet the needs of the segment compared to competitors.
* Segment Evaluation Plot:
  + Plotting: Each segment is represented as a circle on the plot. The x-axis shows segment attractiveness, and the y-axis shows organizational competitiveness.
  + Bubble Size: Indicates an additional criterion like profit potential or another relevant metric.

**Procedure:**

1. Criteria Selection: Define and weight criteria for segment attractiveness and organizational competitiveness.
2. Scoring: Rate each segment on these criteria, multiply by the weights, and sum to obtain scores.
3. Plotting: Use these scores to place segments on the matrix, and adjust bubble sizes to reflect additional metrics.

**PYTHON CODE REPLICATION FOR STEP 8:**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

mcdonalds = pd.read\_csv('/content/drive/MyDrive/mcdonalds.csv')

#convert visit frequency to numeric

visit\_frequency\_mapping = {

    "Never": 0,

    "Once a week": 1,

    "Once a month": 2,

    "Once a year": 3,

    "More than once a week":4,

    "Every three months":5

}

mcdonalds['VisitFrequencyNumeric'] = mcdonalds['VisitFrequency'].map(visit\_frequency\_mapping)

print(mcdonalds['VisitFrequencyNumeric'].head())

# Assuming k4 is a column in the dataset representing the segment membership

visit = mcdonalds.groupby('k4')['VisitFrequencyNumeric'].mean()

like = mcdonalds.groupby('k4')['Like.n'].mean()

mcdonalds['Female'] = (mcdonalds['Gender'] == 'Female').astype(int)

female = mcdonalds.groupby('k4')['Female'].mean()

plt.figure(figsize=(10, 6))

bubble\_size = 10 \* female

plt.scatter(visit, like, s=bubble\_size, alpha=0.5)

# Add text labels for each segment

for i in range(len(visit)):

    plt.text(visit.iloc[i], like.iloc[i], str(visit.index[i]+1), fontsize=12, ha='center', va='center')

plt.xlim(2, 4.5)

plt.ylim(-3, 3)

plt.xlabel('Visit Frequency')

plt.ylabel('Liking (Like.n)')

plt.title('Segment Evaluation Plot')

plt.show()