# **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 significant long-term commitment and investment from organizations. It involves developing new products, modifying existing ones, and changing pricing and distribution channels. This can impact the organization's internal structure, requiring it to organize around market segments.

Deciding to use market segmentation implies the need for continuous and significant investments in research, surveys, focus groups, package designs, and advertisements. It's crucial to ensure that the anticipated increase in sales justifies these costs. The decision 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.

Challenges in Market Segmentation:

Senior Management:

* Lack of leadership and involvement.
* Insufficient resource allocation.

Organizational Culture:

* Resistance to change, lack of market orientation, poor communication.
* Lack of training and understanding in market segmentation.
* Absence of formal marketing function or qualified experts.

Operational Challenges:

* Limited financial resources and inability to make structural changes.
* Process-related issues like unclear objectives, poor planning, and time pressure.
* Management's reluctance to use complex techniques.

**STEP 2:**

**Segment Evaluation Criteria**

In market segmentation analysis, two sets of criteria are established: knock-out criteria and attractiveness criteria. Knock-out criteria filter out segments that don't qualify, while attractiveness criteria assess the appeal of qualifying segments. The segmentation team selects and weighs these criteria to determine overall segment appeal. Knock-out criteria include substantiality, measurability, accessibility, homogeneity, distinctiveness, size, and organizational capability. Attractiveness criteria evaluate appeal based on various factors. The team selects around six criteria, assigning each a weight to indicate its importance. The segment evaluation plot, where segment attractiveness is plotted against organizational competitiveness, is a commonly used approach. Criteria for these evaluations should be discussed and agreed upon by a diverse segmentation team representing various organizational units.

**STEP 3:**

**Segmentation Variables**

The basis for market segmentation is empirical data, utilized for creating market segments and providing detailed descriptions. Common sense segmentation uses single characteristics like gender, while data-driven segmentation uses multiple variables. High-quality data is essential for accurate segmentation, and it can come from surveys, observations, and experimental studies, but should ideally reflect actual 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 analyzing segmentation variables across market segments to gain a better understanding. This process occurs in Step 2 (identifying the ideal target segment), Step 3 (collecting data), and Step 7 (describing segments). Effective segment description is crucial for crafting a targeted marketing mix based on demographic, psychographic, and behavioral data. Visualizations like stacked bar charts and mosaic plots are used to compare descriptor variables across segments.

**Metric Descriptor Variables**

Visualizations for metric variables like income and age can reveal associations with segment membership. For example, mosaic plots can show income distribution across different segments, and analyzing moral obligation scores can highlight how different segments value environmental behaviors.

**Key Insights from Visualizations**

Gender Distribution: Gender balance observed across market segments.

Income Levels: Higher income linked to cultural travel motives; lower income linked to budget-friendly options.

Environmental Morality: Strong environmental motivations associated with higher moral obligation scores.

**Metric Descriptor Variables and Conditional Plots**

Visualizing Segment Differences:

- Using R packages like lattice and ggplot2 helps compare metrics like age and moral obligation across different segments.

- Box-and-whisker plots reflect minimal age differences but significant variations in moral obligation between segments.

- Statistical tests like ANOVA and Kruskal-Wallis reveal notable differences in means and medians across segments.

Pairwise Comparisons:

- Comparisons highlight segments 5 and 6 as having significantly higher moral obligation scores compared to others.

Segment Level Stability Across Solutions (SLSA) Plot:

- SLSA Plot visually represents stability and moral obligation levels across segments over various solutions, highlighting segments with high moral obligation consistently.

**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).
* **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.
* **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**

**Purpose:**

**-** CART is used for predicting binary or categorical dependent variables.

- Trees allow for variable selection, ease of interpretation, and interaction effects, but may be unstable with small data changes.

Implementation in R:

- The rpart and partykit packages are commonly used for tree construction.

- Use ctree() for binary dependent variables:

library("partykit")

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

- Use ctree() for categorical dependent variables with more than two categories:

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

Model Output:

- Trees are visualized with nodes from splits and terminal nodes for final predictions.

Model Evaluation:

- Evaluate tree complexity and proportion of correct predictions in terminal nodes.

**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.

**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.

The matrix has two axes:

- Segment Attractiveness (x-axis): Reflects the segment's desirability based on growth potential, profitability, and fit with organizational goals.

- Organizational Competitiveness (y-axis): Measures how well the organization can meet segment needs compared to competitors. Each segment is represented as a circle on the plot, with bubble size indicating additional criteria.

**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()