Market Segmentation Analysis

Sai Sandeep

Step 1: Deciding (Not) to Segment

- Segmentation is a long-term strategy, not something to try and abandon quickly.
- Only segment if the expected benefits outweigh the cost and effort involved (like new products, messaging, packaging, etc.).
- Leadership support is essential without senior buy-in and resources, the strategy may fail.
- Company culture matters openness to change and collaboration across departments is needed.
- Missing skills or poor planning can become major barriers to success.

Step 2: Specifying the Ideal Target Segment

4.1 Segment Evaluation Criteria

- Decide who the ideal customer is before choosing segments.
- Use two types of criteria:
    - Knock-out criteria: Must-have traits (e.g., size, reachability, internal similarity).
    - Attractiveness criteria: Help compare and rank viable segments based on factors like fit, growth, and profit potential.

4.2 Knock-Out Criteria

- Segments must be:
    - Big enough to justify targeting,
    - Internally similar and externally different,
    - Reachable and identifiable,
    - Aligned with the company's strengths.

4.3 Attractiveness Criteria

- Used to score and rank segments (not a yes/no filter).
- Segments are evaluated on how well they meet selected factors (e.g., profitability, fit, size, risk).

4.4 Implementing a Structured Process

- Use a segment evaluation plot to compare attractiveness vs. company strength.
- Select and weight 5–6 key evaluation criteria.
- Involve different departments to ensure alignment and diverse input.
- Set evaluation criteria early, so the right data is collected and used later.

Step 3: Collecting Data

5.1 Segmentation Variables

- Segmentation variables divide the market (e.g., behavior or needs).
- Descriptor variables describe segments (e.g., age, income).
- Good data quality is essential for accurate segmentation.
- Use data from surveys, purchases, observations, or experiments — behavior data is often more reliable.

5.2 Segmentation Criteria (Types)

5.2.1 Geographic

- Based on location (e.g., region, country).
- Easy to apply but doesn't always reflect true customer needs.

5.2.2 Socio-Demographic

- Based on age, gender, income, etc.
- Easy to collect but explains only a small part of consumer behavior.

5.2.3 Psychographic

- Based on values, interests, beliefs, or motivations.
- More powerful but harder to measure and segment accurately.

5.2.4 Behavioral

- Based on actual behavior (e.g., purchases, frequency, spending).

- Highly useful but data may be hard to get for new customers.

5.3 Data from Survey Studies

5.3.1 Choice of Variables

- Include only necessary variables — avoid irrelevant or duplicate ones.
- Use qualitative research to guide question development.

5.3.2 Response Options

- Binary and metric response types are most suitable for segmentation.
- Ordinal data (e.g., Likert scales) is less precise unless handled carefully.
- Slider scales (visual analogue) are a good modern alternative.

5.3.3 Response Styles

- Watch out for biases like always agreeing or choosing extremes.
- These can distort the segments, so use neutral wording and double-check results.

5.3.4 Sample Size

- Too small a sample leads to poor results.
- General rules:
  - At least 60× number of variables for standard cases.
  - 100× number of variables for best accuracy.
- Clean, relevant, and well-structured data improves outcomes more than sheer size.

5.4 Data from Internal Sources

- Includes sales records, loyalty data, online purchases, etc.
- Reflects real behavior and requires no extra effort to collect.
- However, it often only covers existing customers — not future prospects.

5.5 Data from Experimental Studies

- Comes from lab/field tests (e.g., ad responses, choice experiments).
- Helps segment based on how people react to different features or messages.
- Great for understanding preference and decision-making.

Step 5: Extracting Segments

Step 5 involves extracting market segments using three techniques: k-means analysis, finite mixtures of binary distributions, and finite mixtures of regressions.

Using k-Means

We calculate segmentations for two to eight segments using k-means with ten random restarts. This range is chosen because the optimal number of segments is not known beforehand. The goal is to identify segments with similar consumers while ensuring they differ from other segments. A scree plot is used but doesn't provide clear guidance, so we turn to stability-based data structure analysis to assess the reliability of each segmentation.

Stability-Based Data Structure Analysis

Global stability measures how consistent the segmentation solution is across repeated calculations. The four-segment solution is the most stable, providing a good balance between stability and differentiation. Adding more segments reduces stability, suggesting fewer segments may be more reliable.

Segment Level Stability Analysis (SLSA)

We also evaluate how segment memberships change when the number of segments increases. The SLSA plot shows that while segments 2, 3, and 4 are stable, segment 1 is unstable and splits across different solutions.

Final Segmentation Choice

The four-segment solution is selected for its stability, despite the instability of segment 1. Segment-level stability within solutions (SLSW) shows that segment 1 is the least stable, while segment 3 is the most stable.

Latent Class Analysis Using a Finite Mixture of Binary Distributions

We perform latent class analysis with a finite mixture of binary distributions, maximizing likelihood to extract segments, unlike k-means, which minimizes squared Euclidean distance. The stepFlexmix() function is used to extract two to eight segments with ten random restarts of the EM algorithm.

Choosing the Number of Segments

Information criteria (AIC, BIC, and ICL) are plotted for different segment counts. The values decrease sharply until four segments, after which they flatten. While ICL and BIC suggest seven segments, the AIC continues to drop beyond that. The flattening after four segments suggests a four-segment solution may be optimal.

Comparing with k-Means Solution

A cross-tabulation of the four-segment mixture model and k-means solutions shows that the segments from both methods are similar, particularly for stable segments, despite using different techniques.

Initialisation and Log-Likelihood Comparison

We compare the log-likelihood values of two initializations: random restarts and k-means initialization. The results are almost identical, suggesting that both methods are close to a global optimum and provide reliable segmentations.

In summary, both the mixture model and k-means solutions provide valuable, similar insights, reinforcing the confidence in the chosen segmentation.

Using Mixtures of Regression Models

We segment consumers based on how their perceptions of McDonald's influence their love or hate for the brand, allowing targeted marketing to improve love and reduce hate.

The dependent variable is the degree of liking McDonald's, converted from an 11-point ordinal scale. We use finite mixtures of linear regression models (latent class regressions) with the EM algorithm, fitting two segments.

Handling Ordinal Data Challenges

Since the dependent variable is ordinal, extracting many segments can lead to degenerate solutions. To avoid this, we restrict the model to two components.

Assessing Regression Models

The model shows that Segment 1 values "YUMMY," "FAST," and "TASTY" perceptions, while Segment 2 cares about "CONVENIENT," "HEALTHY," and "YUMMY" perceptions. Segment 1 focuses on taste and value, while Segment 2 requires a health-focused message.

Regression Coefficient Comparison

A plot compares coefficients, with significant values shaded. Segment 1 should be targeted with taste and cost-related messages, while Segment 2 requires health-oriented messaging.