02_Market segmentation

Kaveripakam Jayaprakash Jayashree
14/06/2025

Step 4: Digging into the McDonald's Data

Before beginning segmentation, the McDonald's dataset is analyzed to understand general consumer perceptions. The dataset includes 1,453 responses across 15 variables: 11 perception-based (e.g., "yummy," "cheap," "fast") and 4 demographic/behavioral (Age, Gender, VisitFrequency, Like). Since the perception variables are coded as "Yes" or "No," they are converted into binary format (Yes = 1, No = 0) for analysis. The means of these binary values offer insights into overall opinions—for instance, 91% of respondents see McDonald's as convenient, while only 9% associate it with being spicy.

To reduce complexity and extract meaningful patterns, Principal Component Analysis (PCA) is used. It reveals that the first two components explain nearly half the variance in consumer responses. The second component largely captures a price perception dimension (expensive vs. cheap), while other components represent emotional groupings like positive impressions (e.g., "tasty," "fast") and negative impressions (e.g., "greasy," "disgusting"). The resulting perceptual map suggests three major consumer mindsets: value-seeking, price-conscious, and neutral, laying the foundation for data-driven segmentation in the next step.

Step 5: Applying Clustering Methods

5.1 Using k-Means Clustering

In this step, clustering is used to identify distinct customer segments. k-means clustering is applied with cluster counts ranging from two to eight. A scree plot is generated to assess the optimal number of clusters, but no clear "elbow" suggests an ideal choice. To overcome this, a more robust technique—stability-based analysis—is used. By repeatedly running k-means on resampled data, a stability plot shows that four segments offer the best balance between differentiation and consistency. Visual tools like gorge plots and Segment Level Stability Across Solutions (SLSA) confirm that segments 2, 3, and 4 are highly stable, while segment 1 is unstable.

Based on these insights, the four-segment solution is selected. Further stability analysis (SLSW plot) again reveals segment 3 as the most consistent and segment 1 as the least. This means McDonald's can confidently target well-defined groups like segments 2, 3, and 4, while being cautious with segment 1, which may vary across different clustering runs. Understanding this stability is crucial for making reliable marketing decisions.

5.2 Using Mixtures of Distributions (Latent Class Analysis)

An alternative method, Latent Class Analysis (LCA), is employed using binary mixture models. This approach optimizes likelihood rather than distances. The Expectation-Maximization (EM) algorithm is used with multiple random starts. The model is evaluated using AIC, BIC, and ICL criteria. Although AIC favors more segments and BIC/ICL suggest seven, a visual inspection of the criteria curves indicates four segments as a reasonable compromise. The four-segment LCA model is then compared with the k-means clusters through cross-tabulation, showing strong overlap—especially in the stable segments (e.g., k-means segment 3 aligns closely with LCA segment 4).

To confirm reliability, the LCA is re-run using initial segment assignments from k-means. Results are nearly identical to the original random-start model, and similar log-likelihood values confirm that both approaches are converging toward a global or near-optimal solution. This agreement boosts confidence in the extracted segments and shows that even with different algorithms, consistent and actionable customer insights emerge.

5.3 Using Mixtures of Regression Models

In this model, segmentation is based on how different perceptions influence customer liking or disliking McDonald's. A latent class regression model is created with the dependent variable being a customer's rating of McDonald's (from -5 = "I hate it" to +5 = "I love it") and the independent variables being perception-based traits like "yummy," "cheap," or "disgusting." This method groups customers based on how their attitudes are shaped by similar drivers.

The analysis yields two key segments. Segment 1 prefers McDonald's when it is perceived as yummy, cheap, fast, and not disgusting, while health-related attributes are not influential. Segment 2, however, values attributes like "healthy," "easy to eat," and "not greasy"—surprisingly, taste factors matter less for them. By visualizing regression coefficients, marketers can easily identify which traits drive positive or negative perceptions. For example, McDonald's might emphasize affordability and taste for Segment 1, but focus on healthy, clean, and lighter options for Segment 2. These distinctions enable highly targeted marketing strategies.

Step 6: Profiling Segments

Now that the four k-means segments are identified, this step investigates how customer perceptions vary across them. A segment profile plot is created, which groups similar attributes using hierarchical clustering (e.g., "tasty" and "yummy" together), making interpretation easier. The percentage of customers within each segment who associate McDonald's with each perception is plotted, with significant deviations from the average marked as key traits.

This plot reveals segment-specific views: Segment 1 sees McDonald's as cheap and greasy; Segment 2 finds it expensive and disgusting; Segment 3 thinks it's tasty and yummy but also expensive; Segment 4 sees it as tasty, cheap, healthy, and yummy. Additionally, a segment separation plot uses PCA to show how segments differ visually in perceptual space. For instance, both Segments 2 and 3 view McDonald's as expensive, but Segment 3 still enjoys the food, while Segment 2 does not. These insights provide a nuanced understanding of customer groups and serve as a foundation for targeted communication strategies.

Step 7: Describing Segments

This step adds demographic and behavioral context to the segments. Since the dataset includes limited descriptor variables, initial analysis uses mosaic plots to relate segment membership with overall liking of McDonald’s, gender, and visit frequency. Results show that Segment 4 tends to like McDonald’s the most, while Segment 2 dislikes it. Segment 2 also contains more males, while Segment 4 has a higher proportion of females.

Box plots comparing age reveal that Segment 3 (who like the taste but think it's pricey) are younger than others. A conditional inference tree is used to predict Segment 3 membership based on age, gender, visit frequency, and liking score. The tree identifies that younger people who either like or visit McDonald’s frequently—even if they don’t love it—are most likely to belong to Segment 3. Interestingly, gender doesn’t significantly affect this segment's profile. While these insights are useful, the analysis suggests that richer descriptor variables (e.g., lifestyle, habits, media exposure) are needed for more actionable marketing segmentation.

Step 8: Selecting Target Segments

To decide which segments to target, a segment evaluation plot is used. This plot considers visit frequency (x-axis), liking score (y-axis), and gender (bubble size = % female) to assess each segment’s attractiveness. Segments 3 and 4 stand out—they are both enthusiastic about McDonald’s and visit often, making them prime targets for retention and growth.

Segment 2, in contrast, is the least appealing—customers here dislike McDonald’s and visit rarely. Segment 1 falls in the middle: although customers in this group view McDonald’s as expensive and feel neutral or slightly negative, they still visit somewhat regularly. This presents an opportunity: with the right marketing tactics (such as pricing strategies or new offerings), this segment might be persuaded to become more loyal. These findings guide decisions for focusing marketing resources effectively in the next step.

Step 9: Customizing the Marketing Mix

With target segments identified, McDonald's can now tailor its marketing mix. For example, to engage Segment 3 (young consumers who like the taste but find it expensive), a special "McSuperBudget" menu could be launched to align with their price sensitivity. This lower-cost menu could offer simplified or smaller portions (Product) to avoid cannibalizing premium offerings. Promotion should target youth platforms like social media and influencer networks to ensure relevance and reach.

While Place remains McDonald's restaurants, operational tweaks—like a separate counter or digital ordering option for budget items—can help serve this segment efficiently without diluting the brand's premium image. By customizing the 4 Ps—Product, Price, Promotion, and Place—McDonald's can better meet the expectations of each segment and boost brand loyalty in a data-driven, targeted way.

Github link:

https://github.com/Jayashree24092004/Newfolder.git