Summary on Market Segmentation Steps

Saurav Kumar Dwivedi

Step 1: Deciding (not) to Segment:

Deciding whether to segment or not is an important step and the points below should be kept in mind before proceeding to Segment:

1.1 Implications of Committing to Market Segmentation

- Segmentation process is not a short term process, it takes a long time. • It is not always a best to perform segmentation as it involves costs of performing the research, fielding surveys, and focus groups, designing multiple packages, and designing multiple advertisements and communication messages.
- It is viable to segment only if the increase in sales of the products in question, justify the time and resource investment to it.
- As new products and marketing needs to be developed, the whole company dynamics change, which should be kept in mind.

1.2 Implementation Barriers

- Several Obstacles are there including lack of Leadership in higher management and lack of commitment.
- Company Culture is not market oriented.
- Lack of qualified marketing expert or lack of training among employees and higher management related to market segmentation.

1.3 Structured Approach

- To perform segmentation a company needs to have long term goal, open to new ideas, market based culture and open to make significant changes within.
- A dedicated team including market experts, data analysts should be present.

Step 2: Specifying the Ideal Target Segment :

The selection of target Segment depends on several factors:

2.1 Segment Evaluation Criteria

- There are several criteria based on which a target or some target segments are chosen suck as knock-out criteria, attractiveness criteria.

- The knock-out criteria decides whether a segment should be accessed using attractiveness criteria or it should be rejected.

2.2 Knock-Out Criteria

- A segment is selected or knocked-out based in several factors e.g, whether the segments are homogenous, the different segments should be distinct, segments should be large, Identifiable segments, it should be reachable and finally Segment should rematching the strengths of organisation

- Senior management and segmentation team should be working on them.

2.3 Attractiveness Criteria

- There are several attractiveness criteria which gauges the acceptance of a segment and they are used depending on specific situation. These criteria are not binary in nature. They are rated out of some total points and then chosen accordingly.

2.4 Implementing a Structured Process • It is always beneficial to have a structured approach towards segmentation. One of the most popular approach is to use an Attractiveness vs Competitiveness plot called segment evaluation plot.

- A team should there to complete the process. The attractiveness criteria can't be easily chosen at this step but in doing so it becomes easier to evaluate the criteria at step 8. At the end os this step the team should have about six such attractiveness criteria. Then they are proposed to the advisory committee.

2.5 Structured Approach

- As far as the tasks go, the segmentation team discusses the knock-out criteria which helps eliminate segments in step 8.

- The team members now access various segment attractiveness criteria and assign weightage to them, which then crosschecked with other team members. Upon agreement they are proposed to advisory committee.

Step 3: Collecting Data:

Data collection and variable choice is performed here:

3.1 Segmentation Variables

- The collected empirical data can be used for both Common-sense and Data Driven segmentation. The variables on the basis of which we perform segmentation are called Segmentation Variables and the variables describing the segments called Descriptor Variables.

- In common-sense segmentation, the segmentation variable is a single variable chosen from the empirical dataset on which segmentation is performed.

- In Data-Driven segmentation, the segmentation variable can be a set of variables which help to find out the natural or artificial segmentations present. In both cases quality of empirical dataset is critical.

3.2 Segmentation Criteria

- The segmentation is also dependent on Segmentation Criteria. The term is used in a broader sense compared to segmentation variable as the criteria can depend on various factors.

- Geographic Segmentation where the location information is used as the basis of segmentation, which is being used from the beginning. Customers are assigned to their geographic location and segmented on that, which helps in categorising different customers in different regions. News papers, TV, radio stations, Tourism industries may benefit from it.

- Socio-Demographic Segmentation uses customer age, income, gender and education for segmentation. Luxury goods, cars and property like real estate industry can be benefited from such segmentations.

- In Psychographic Segmentation customer's beliefs, culture are taken as segmentation criteria. This is a more complex kind of segmentation, but much more reflective of customer behaviour. Tourism industry might help from such segmentation as people might want to explore culturally rich locations more.

- Behavioural Segmentation uses prior behaviours like amount spend, experience with a product, etc. as a criteria. In such cases customer behaviours can be taken as a segmentation criteria which is advantageous.

3.3 Data from Survey Studies

- Data Can be easily generated from surveys for segmentation purpose but they usually has biases. To prevent such biases, we have to choose segmentation variables carefully and must exclude unnecessary variables which divert the segmentation away from the intended information. They are called noisy variables. To avoid that, survey questions must be unique and unnecessary questions must be avoided.

- The response options for the surveys should follow some rules such that segmentation can be unbiased. Binary response options are a good choice. Also, having nominal options where the respondents choose an option from unordered option can be a good way. Allowing the respondents to specify a number, is also a good option as it allows fro statistical analysis. Thus binary or metric method is useful rather that ordinal options where respondents choose their agreements with a series of statements.

- Response Styles may also affect segmentation by including biases. They tend to use neutral or extreme options in surveys. Such responses should be removed for unbiased segmentation, as segmentation algorithms can't capture respondent's belief and also response style.

- Sample Size should be enough to be able to correctly perform segmentation in any number of segmentation variable. A rule of thumb is to take 2^p as the size of sample where p is the number of segmentation variables. If the segments are equally sized clusters sample size should be 10 times the number of segmentation variables. In case of unequal sized clusters, the smaller cluster should have 10p number of samples. But, increasing the sample size is always better for segmentation.

- Therefore, in general the data should contain all necessary items and no unnecessary items, no correlated item and only high quality responses. Only binary or metric responses should be chosen and they should be free from response styles. Sample size should be sufficient.

3.4 Data from Internal Sources

- The data can also be obtained internally from the organisations. Like, realtime customer behaviour from retail stores. They are actual behaviours and not user responses. But, in such case existing customer data is repeated and new customer behaviour is often missing which might have helped in segmentation.

3.5 Data from Experimental Studies

- Data can also be generated synthetically in laboratory or field experiments from user responses from advertisements.

3.6 Structured Approach

- To make all of this happen, a structured approach should be taken within the organisation. An approach would be: Discussing consumer characteristics which should be used as segmentation variables and also to extract segment details. Determine a way to collect data such that segmentation variable and descriptor variables can be identified. And lastly, design a method to collect error free and unbiased data, and actually collect the data.

Step 5: Extracting Segments:

5.1 Data Structure Analysis

- Segment Extraction is an exploratory task, cross-checking and validating is not possible as the organisation can't have multiple segmentation strategies together. Thus, validation in segmentation means slightly tweaking the algorithms or data structures. This represents stability of solutions in repeated calculations and called Stability-Based Data Structure Analysis.

- Stability based data structure analysis helps to find inherent structures in data which forms clusters and also suitable number of clusters. • There are four types of data structure analysis: cluster indices, gorge plots, global stability analysis, and segment level stability analysis.

5.2 Cluster Indices

- To determine important informations like number of clusters, cluster indices in a very good approach. There are two types of cluster indices:
  - Internal Cluster Indices, External Cluster Indices.

- Internal cluster indices are calculated with only one market segment solution and the contained information used as guidance. They ask how separated the clusters are and how compact each segments are.

- To measure the compactness of clusters one way is to use sum of withincluster distances W_k for a segmentation solution with k segments in a centroid based approach.

$$W_k = \sum_{h=1}^{k} \sum_{x \in sh} d(x - ch)$$

- Now, using K-means algorithm, we plot the Elbow curve to determine optimal number of clusters. Several other internal cluster indices are there.

- For the case of External Cluster Indices, they not only depend on the current cluster but also need additional cluster informations. The true segment structure – if known – is the most valuable additional piece of information. But the true segment structure of the data is typically only known for artificially generated data. The true segment structure of consumer data is never known.

- When working with consumer data, the market segmentation solution obtained using a repeated calculation can be used as additional, external information. • We basically take another segmentation solution and find similarities. There is label switching problem when solution stays even after labels change. One way around the problem of label switching is to focus on whether pairs of consumers are assigned to the same segments repeatedly.

- Selecting any two consumers, the following four situations can occur when comparing two market segmentation solutions P1 and P2:

a: Both consumers are assigned to the same segment twice. b: The two consumers are in the same segment in P1, but not in P2.
c: The two consumers are in the same segment in P2, but not in P1.
d: The two consumers are assigned to different market segments twice.

The index proposed by Jaccard (1912) is based on this observation, but uses only a, b and c while dropping d: J = a/(a+b+c). There are several other external indices.

5.3 Gorge Plots

- Another way to see how well separated the segments are, is to use similarity formula:

$$S_{ih} = \frac{e^{-d_{ih}^{y}}}{\sum_{l=1}^{K} e^{-d_{il}^{y}}}$$

Similarity values can be visualised using Gorge plots. High similarity values indicate that a consumer is very close to the centroid of the market segment. Low similarity values indicate that the consumer is far away from the centroid.

- If the similarity values are the result of model-based segment extraction methods, high similarity values indicate that a consumer has a high probability of being a member of the market segment. Low similarity values indicate low probability of segment membership.

- If the data is natural, there will be very high and very low values and the poor looks like a gorge with peak towards very high and very low.

5.4 Global Stability Analysis

- Another alternative way is Global Stability Analysis where several solutions are generated using resampling. Then stability of such solutions are checked and the most stable one is chosen.

- In the best case if natural segments exists in data then it is easily to detect.

- In the worst case if the data is completely unstructured, then the analyst must inform that and produce managerially useful segments. If segments are constructive, then the analyst must produce interesting segments and determine the best segment.

- In case of a middle option between the worst case and the best case scenario, consumer data can lack distinct, well-separated natural clusters, while not being entirely unstructured, the existing structure can be leveraged to extract artificially created segments that re-emerge across repeated calculations. This case is referred to as reproducible segmentation.

- Global Stability Analysis determines which concept to apply. The results from global stability analysis also assist in determining the most suitable number of segments to extract from the data.

- For higher-dimensional data, where it is impossible to plot the data to determine its structure, it is unavoidable to conduct stability analysis to gain insight into the nature of the market segmentation solution.

5.5 Segment Level Stability Analysis

- Global Stability does not always choose the best segment, thus it is recommended to check segment level stability analysis. In a lot of discarded solutions we might find useful segment informations.

- The first criteria for segment level analysis is a similar concept to Global Stability analysis is used in segment level stability within solutions (SLSW), the difference being the stability, computed at segment level, allowing the detection of one highly stable segment in a segmentation solution where several or even all other segments are unstable.

- Segment level stability within solutions (SLSW) measures how often a market segment with the same characteristics is identified across a number of repeated calculations of segmentation solutions with the *same* number of segments. • The second criteria is, segment level stability across solutions (SLSA). The idea is to find if a segment reoccurs in different segmentation solutions having different number of segments.

- High values of segment level stability across solutions (SLSA) means market segments occurring naturally in the data, rather than being artificially created. Natural segments are more attractive to organisations because they actually exist, and no managerial judgement is needed in the artificial construction of segments.

- Segment level stability across solutions (SLSA), can be calculated in combination with any algorithm which extracts segments. For hierarchal clustering, the stability analysis shows a sequence of partitions created. If partitioning methods (k-means, k-medians, etc.), segmentation solutions are determined separately for each number of segments k

- The segment level stability plot is based on visualisation and for the measure of entropy Shanons entropy can be used. Entropy is defined as $\sum(-p_i log p_i)$ and measures the uncertainty in a distribution. Numerical stability SLSA(Sil) of segment l in the segmentation solution with k_i segments is defined by:

$$SLS_A(S_{li}) = 1 - \frac{\sum_{j=1}^{K_{i-1}} P_j log P_j}{\log (k_{i-1})}$$

0 being minimum and 1 being maximum stability.

5.6 Structured Approach

- Select and use suitable segment extraction methods depending on the data to group customers.
- Perform Global and Segment based stability analysis to find the best segmentation solution and promising segments.
- Assess the remaining segments using Knock-out criteria and get the final segments.

****

McDonald Dataset Implementation (Code Conversion)

Github Link-

https://github.com/saurav4695/McDonald_Market_Segmentation