**Market Segmentation Analysis**

**Introduction:-**

Market Segmentation is a vital marketing tool, which involves dividing diverse markets into smaller homogeneous segments based on key consumer attributes. This helps in precise market selection and tailored marketing efforts. Consumers within a segment should share key characteristics while differing from other segments. Approaches such as concentrated and differentiated serves a specific segment. In contrast, an undifferentiated approach considers the entire market to be one, relevant for resource rich situations or new product introduction. Firms rely on segmentation to shape their marketing decisions.

**Step 1:- Deciding (not) to Segment**

**Implications of market segmentation:-**

The decision of market segmentation comes with profound implications. Hence, it is important to understand market segmentation may demand substantial investment and changes for the organization and also it is a long term commitment. Before deciding segmentation analysis, an organization should understand its full implications. Successful implementation results in organizational adaptation - product, pricing, communication and internal structure may change. Once we segment the market, it is advised that we organize around market segments, not product and services. High-level executives should make the decision and must ensure consistent communication and reinforcement across the organization, as it is a long term commitment.

**Implementation barriers :-**

Market segmentation implementation is not straightforward and might involve several barriers. Higher management lacking commitment, understanding and presence of inadequate resources in the organization can result in failure. For successful implementation, it is important to transcend lack of orientation, resistance to change, poor communication and short-term thinking. Also, absence of data management and marketing expertise can result in inefficient execution or failure of market segmentation. In addition, process related issues such as unclear objectives or planning are barriers. Addressing the aforementioned issues is crucial for success. If the barriers persist, then careful consideration of whether or not to continue pursuing segmentation, is advised.

**Step 2:- Specifying the target segment**

**Segment evaluation criteria :-**

While segment evaluation user input is crucial criteria, use involvement should span for various stages, influencing technical aspects. After deciding to segment (step 1), the organization progresses conceptually in step 2, selecting segmentation evaluation criteria. Knock-out criteria is non-negotiable, it defines essential segment features. Attractiveness criteria evaluates remaining segments meeting knock- out criteria.

At places we may find them lumped together, but they serve different purposes. Knock- out criterias are unchanging, while attractiveness criteria offer a range of options. Groups involved in segmentation decide which attractiveness criteria to apply by evaluating their importance. Balancing these criterias shapes the target segment selection (Step 8).

**Knock-Out Criteria :-**

Knock-Out Criteria help us determine if market segments are eligible for further consideration. This criteria involves homogeneity, distinctiveness, alignment with available organizational resources, identifiability and reachability. Homogeneity requires segment members to be similar, distinctiveness mandates clear differences from other segments. Understanding aforementioned criteria are important for higher management, advisors and people actually doing segmentation.

**Attractiveness Criteria :-**

Attractiveness Criteria complements knock-out criteria. Attractiveness isn’t binary; segments are rated based on different criteria. Attractiveness criteria helps in target segment selection in step 8.

**Step 3: Collecting Data :-**

**Segmentation Variables :-**

Segmentation variables are crucial for commonsense and data driven market segmentation. In commonsense segmentation, a single characteristic such as gender splits the sample into segments. Descriptor variables such as demographic and behavioral, provide segment details. Data driven segmentation uses multiple variables to create or identify segments. Highly accurate and empirical data is crucial, ensuring appropriate segment assignment. For efficient solutions, accurate data is vital irrespective of common sense or data driven approach. Data sources include surveys, observations or experiments. Accurate data reflecting consumer behavior is important to ensure successful market segmentation analysis.

**Segmentation Criteria :-**

The choice of segmentation criteria is of vital importance before data collection or extraction. The term segmentation criterion encompasses the nature of information used for segmentation. Mostly used criteria include geographic, socio-demographic, psychographic and behavioral factors. In addition, factors such as preferences, profitability and consumer interaction influences segmentation.With various criteria available, simplicity is recommended.

**Geographic Segmentation:-**

Geographic segmentation uses consumer’s location to form market segments. It is useful when language or location-specific factors affect consumer behavior. It enables precise targeting but may fail to observe other relevant characteristics.

**Socio-demographic :-**

Socio-demographic factors such as age, gender, income and education are useful in industries such as cosmetics and baby products. They offer clear segment identification but may not fully explain product preferences. Values, tastes and preferences are more influential in consumer behavior.

**Psychographic Segmentation :-**

Psychographic segmentation groups people based on beliefs, preferences and interests. It is complex, relying on multiple variables but it gives insights into consumer behavior and preferences.

**Behavioral Segmentation :-**

Behavioral segmentation is by using actual behavior such as purchase frequency and amount spent. It groups people by their actions, providing insights on empirical consumer behaviors.

Data availability is a potential issue with behavioral segmentation.

**Data from survey studies:-**

Most market segmentation analyses are based on survey data. Surveys are cheap and easy to collect, convenient for any organization. But survey data in contrast to data obtained from actual behavior can be influenced by various biases, potentially reducing quality of segmentation analysis. There are several factors which need to be considered when using survey data, these include choice of variables, response options, response styles and sample size.

**Data from internal sources:-**

In this day and age, organizations have access to a significant amount of internal data which can be utilized for market segmentation analysis. Data such as airline booking and online purchase can be used for segmentation. This data reflects actual consumer behavior avoiding biases inherent to self-reported data.

Current data can be biased towards existing consumers and may lack insights for potential future customers with different consumer patterns.

**Data from experimental studies:-**

Experimental data such as data obtained from lab experiments can be a valuable source for segmentation. Experimental data can provide insights into responses to ads and preferences for products with various attribute combinations, offering insights into customer patterns for segmentation purposes.

**Step 4: Exploring Data**

**First glimpse at the data:-**

Once we collect the data, exploratory data analysis cleans and if necessary preprocess the data for market segmentation. This exploration also gives us insights into suitable algorithms for extracting meaningful market segments. Here the travel motive dataset is used as an example, columns such as gender, Age, Income and Income 2 are analyzed. Missing values, denoted as NAs, are identified. In the given dataset, 66 respondents have missing income data. Besides, R code is provided to showcase data inspection, summary generation, and basic statistics, to help understand and prepare the dataset for market segmentation analysis.

**Data Cleaning:-**

Data cleaning is essential for reliable data analysis. It involves checking data consistency and accuracy. In R, implausible values such as age outliers can be identified. Categorical values such as gender, must only contain permissible values. In the given case, income categories were out-of-order. Reordering them involves creating a helper variable, recording levels and converting variables into ordered factors. Verification can be done through cross-tabulation to ensure correctness. Using code for data cleaning documents the process. Cleaned Data in R can be saved and loaded for future Sessions.

**Descriptive Analysis:-**

Descriptive analysis is important to understand the data, before complex analysis. In R, we can use the “summary()” method to get numeric summaries for variables. In addition to the summary() method, we can use histograms, box plots and scatter plots for numeric data. Bar plots visualize categorical variables and mosaic plots demonstrate association between multiple categorical variables. Histograms show data distribution and can be customized by specifying the number of bins.

Box-and-whisker plots provide a summary of metric variables, including minimum, median and maximum. Graphical representation, such as dot charts, can reveal data patterns, providing insights for further analysis.

**Pre-Processing**

**Categorical Variables :-**

Categorical variables often need pro-processing, including merging categories and converting them to numerical values. Merging categories can help when working with too many original categories. For instance, merging low-frequency income into broader ones. Converting categorical variables into numerical ones is possible if distances between ordinal scale points are approximately equal. Ordinal scales such as Likert scales (used in surveys) can be treated as numeric if distance between their options are considered equal. It is of vital importance to consider implications of chosen response options and their impact on data quality. Binary variables are often preferred for their simplicity.

**Numeric Variables :-**

Standardization is a process used to bring variables onto a common scale, ensuring they have comparable influence in analysis. By standardizing variables their values are transformed so that they have mean of zero and standard deviation of one. This process allows variables with different ranges to be treated equally. In R we can standardize variables using the ‘scale()’ function.

**Principal Component Analysis:-**

Principal component analysis (PCA) is a method used to transform multivariate data into a new set of uncorrelated variables called principal components. These components are sorted by importance, with the first component containing most variability. PCA maintains the relative position of observations while reducing dimensionality. It operates on covariance and correlation matrix of variables. PCA is used to project high dimensional data into lower dimensions for visualization. Using PCA for segmentation variables is discouraged, as it can cause loss of valuable information. Instead, it can help in exploring data and identifying redundant variables.

**Step 5: Extracting Segments**

**Grouping Consumers:-**

Segmentation is data driven and exploratory. Customer data is often highly unstructured and lacks clear groups. Various clustering methods can help shape segmentation solutions. Algorithms may impose structures. No one-size-fits-all algorithm exists. We can consider segment characteristics and data while choosing the methods. Factors such as data size and specific structure influence the choice. In addition, Directly and Indirectly observable characteristics impact model selection. Asymmetry in binary variables is recommended when extracting segments.

**Distance-Based Methods:-**

The goal is to group tourists based on vacation activity preferences. A dataset with seven people’s vacation preferences is considered, including BEACH, ACTION, and CULTURE activities.

Anna and Bill share identical profiles, hence they belong to the same segment. Michael stands out as the only one disinterested in the beach, distinguishing him from others. For implementing this grouping, a mathematical measure of similarity and dissimilarity, often called a distance measure is an important consideration.

**Model-Based Methods:-**

Model based methods provide alternatives to distance based approaches in market segmentation.Model based methods assume that the market segments possess certain characteristics : members share unique characteristics and specific size. Model based methods empirical data to determine segment size and specific characteristics. These methods involve finite mixture models, where consumers are assigned to segments based on probabilities. Despite being complex, these models succeed in capturing intricate segment characteristics.

**Algorithms with integrated variable selection:-**

Most segmentation algorithms assume that all variables contribute equally, but this may not always be true, especially when binary data contains redundant and noisy variables. Traditionally used preprocessing methods can’t filter such variables in binary data. Instead, there are some algorithms which do variable selection during segment extraction. Two methods for binary data are biclustering and Variable Selection Procedure for Clustering Binary Data (VSBD). These and similar approaches ensure appropriate variable selection in binary segmentation problems.

**Data Structure Analysis**

Segmentation analysis is inherently exploratory, which makes validation challenging. Validation mainly refers to checking reliability and stability across repeated calculations with slight data or algorithm modifications, this approach is known as stability based data structure analysis. It helps in determining if natural,distinct and separated market segments exist in the data. If they do, they can be easily revealed, if not we need to explore various solutions. Data structure analysis offers insights into data properties and assists in choosing an appropriate number of segments, using methods such as cluster indices, gorge plots, global stability analysis, and segment-level stability analysis.