Market Segmentation Analysis

1.Team Leader : Pratiksha Gaikwad.

2.Team Member : B Pavan Srinivasarao.

3.Team Member : Himesh Ray.

4.Team Member : Prajakta Kulkarni.

Pratiksha Gaikwad.

Step1

Step2

Step3

Step5

Step 1: Deciding (not) to Segment:

1.1 **Implications of Committing to Market Segmentation**

- The key implication is that the organisation needs to commit to the segmentation strategy on the long term.
- The commitment to market segmentation goes hand in hand with the willingness and ability of the organisation to make substantial changes (McDonald and Dunbar 1995) and investments.
- Potentially required changes include the development of new products, the modification of existing products, changes in pricing and distribution channels used to sell the product, as well as all communications with the market.
- Because of the major implications of such a long-term organisational commitment, the decision to investigate the potential of a market segmentation strategy must be made at the highest executive level, and must be systematically and continuously communicated and reinforced at all organisational levels and across all organisational units

1.2. Implementation Barriers

- The first group of barriers relates to senior management.
- A second group of barriers relates to organisational culture.
- Another potential problem is lack of training.
- Closely linked to these barriers is the lack of a formal marketing function or at least a qualified marketing expert in the organisation.
- Another obstacle may be objective restrictions faced by the organisation, including lack of financial resources, or the inability to make the structural changes required.

Step 2: Specifying the Ideal Target Segment

2.1 Segment Evaluation Criteria

- User input is crucial at multiple stages of market segmentation analysis, not just at the beginning or end.
- Organizations contribute conceptually in Step 2 by determining evaluation criteria for segments, which guide subsequent steps.
- Two sets of segment evaluation criteria are identified:
    a. Essential, non-negotiable features required for targeting segments.
    b. Used to evaluate the relative attractiveness of segments that meet the knockout criteria.
- The literature proposes a variety of segment evaluation criteria, often without distinguishing between knock-out and attractiveness criteria.

- Criteria proposed by different authors include factors such as measurability, substantiality, accessibility, competitive advantage, profitability, and socio-political considerations.
- Knock-out criteria automatically eliminate some segments, while attractiveness criteria are negotiated and applied to determine the overall attractiveness of the remaining segments.

2.2. Knock-Out Criteria

- used to determine if market segments resulting from the market segmentation analysis qualify to be assessed using segment attractiveness criteria.
- The segment must be homogeneous, large enough, distinct, matching the strengths of the organisation, identifiable, reachable.
- Knock-out criteria must be understood by senior management, the segmentation team, and the advisory committee.

2.3 Attractiveness Criteria

- Attractiveness criteria are not binary in nature.
- Segments are not assessed as either complying or not complying with attractiveness criteria.
- The overall attractiveness of a segment, determined by its ratings across all criteria, guides the selection of target segments in Step 8 of the market segmentation analysis.

2.4 Implementing a Structured Process

- Using a structured process for assessing market segments is beneficial, as agreed upon in the literature.
- Evaluates segments based on attractiveness and organizational competitiveness, with criteria determined by the team.
- Criteria must be negotiated and agreed upon, as no standard set fits all organizations.
- A core team proposes criteria, which are reviewed and modified by an advisory committee representing all organizational units.
- Selecting criteria early ensures relevant data collection and simplifies target segment selection.
- Around six attractiveness criteria should be identified and weighted by distributing 100 points among them, with advisory committee approval.

Step 3: Collecting Data

3.1 Segmentation Variables

- **Empirical Data**: Essential for identifying and describing market segments, using either commonsense or data-driven approaches.
- **Segmentation Variables**: In commonsense segmentation, one variable (e.g., gender) splits the sample; in data-driven segmentation, multiple variables identify segments.

- **Descriptor Variables**: Used to detail segments, including socio-demographics and media behavior.
- **Data Quality**: Critical for accurate segment assignment and description, impacting product development, pricing, distribution, and communication strategies.
- **Data Sources**: Empirical data can come from surveys, observations, or experimental studies. Behavior-reflective data is preferred over survey data, which may be unreliable.

3.2 Segmentation Criteria

- Before extracting segments and collecting data, organizations must choose a segmentation criterion, which is broader than a segmentation variable and relates to the type of information used for market segmentation.
- This decision requires market knowledge and cannot be easily outsourced.
- Common criteria include geographic, socio-demographic, psychographic, and behavioral factors.
- Key consumer differences relevant to segmentation are profitability, bargaining power, benefit preferences, barriers to choice, and interaction effects.
- The best practice is to use the simplest, most cost-effective criterion that works for the product or service.

3.2.1 Geographic Segmentation

- Geographic information is often used as a segmentation criterion due to its simplicity and practicality.
- It involves segmenting consumers based on their location of residence, which is useful for tailoring language and marketing strategies, as seen with Austria's tourism board or global companies like Amazon and IKEA.
- The main advantage is the ease of assigning consumers to geographic units and targeting local communication channels. However, geographic segmentation has limitations, as people in the same area might not share other relevant characteristics, like product preferences.
- Despite its drawbacks, geographic segmentation remains valuable, especially in international market studies where segmentation variables need to be meaningful across diverse regions.

3.2.2 Socio-Demographic Segmentation

- Socio-demographic segmentation uses criteria like age, gender, income, and education to create segments.
- These segments are particularly useful in industries like luxury goods, cosmetics, baby products, retirement villages, and tourism, where certain socio-demographic factors closely relate to product preferences.
- The main advantage is the ease of determining segment membership for each consumer.
- In some cases, socio-demographic factors can explain product preferences, such as families with children choosing family vacation spots. However, socio-demographics often do not provide sufficient insight into consumer behavior, explaining only about 5% of the variance

according to Haley (1985). Yankelovich and Meer (2006) argue that values, tastes, and preferences are more influential for segmentation.

3.2.3 Psychographic Segmentation

- Psychographic segmentation groups people based on psychological criteria such as beliefs, interests, preferences, aspirations, or benefits sought.
- Introduced by Haley (1985) as a broad term for all measures of the mind, benefit segmentation and lifestyle segmentation are common approaches.
- Unlike geographic or socio-demographic criteria, psychographic segmentation often requires multiple variables to capture the complexity of consumer motivations.
- This approach is advantageous as it better reflects the reasons behind consumer behavior, such as cultural tourists seeking destinations rich in cultural experiences.
- However, the complexity of determining segment memberships and the reliance on the reliability and validity of psychographic measures are notable challenges.

3.2.4 Behavioural Segmentation

- Behavioural segmentation categorizes individuals based on their observed actions or reported behaviors, such as purchase frequency, spending habits, or brand loyalty.
- Unlike other segmentation methods, it focuses on actual consumer behavior rather than demographic or psychographic traits.
- This approach offers precise insights into consumer preferences and purchasing patterns, making it particularly effective for targeting specific customer segments with tailored marketing strategies.
- However, accessing comprehensive behavioral data can be challenging, especially when including potential customers who haven't previously interacted with the product or brand.

3.3 Data from Survey Studies

- Most market segmentation analyses are based on survey data. Survey data is cheap and easy to collect.

3.3.1Choice of Variables

- Carefully choosing segmentation variables is crucial for effective market segmentation, whether in commonsense or data-driven approaches. In data-driven segmentation, all relevant variables capturing the segmentation criteria must be included, while unnecessary variables should be avoided.
- Including unnecessary variables can lead to respondent fatigue and lower response quality, complicating the segmentation process unnecessarily.
- These "noisy" variables do not contribute essential information and can obscure the correct segmentation solution, affecting algorithms' ability to identify optimal market segments.

- To mitigate this, surveys should focus on essential questions without redundant items, ensuring clarity and relevance in data collection for accurate segmentation outcomes.

3.3.2 Response Options

- In survey research, the choice of answer options directly impacts the data's suitability for segmentation analysis.
- Binary responses, represented as 0s and 1s, are straightforward for distance-based techniques. Nominal variables, where respondents choose one option from an unordered list (like occupation), can be transformed into binary format.
- Metric data, such as age or number of nights stayed, are ideal for statistical analysis due to their measurable distances.
- Ordinal data, common in surveys with multiple ordered response options, lack defined distances between responses, complicating segmentation analysis unless specific assumptions are made.
- Opting for binary or metric responses, when possible and meaningful, minimizes such complexities.
- Visual analogue scales, increasingly used online as slider scales, provide metric-like data and can capture nuanced responses effectively, often outperforming ordinal scales in clarity and precision.

3.3.3 Response Styles

- Survey data can be influenced by response biases, where respondents consistently answer based on a general tendency rather than specific item content, known as response styles.
- These biases, like acquiescence (agreeing with all statements), can skew segmentation results by creating misleadingly homogeneous segments. For instance, a segment indicating high spending on vacation activities may reflect a response style rather than genuine behavior.
- Minimizing response styles during data collection is crucial to ensure accurate market segmentation, requiring careful analysis or exclusion of affected respondents when identifying target segments.

3.3.4 Sample Size

- To ensure accurate market segmentation outcomes, survey data must adhere to several critical standards.
- Firstly, it should include all necessary variables while excluding any irrelevant or redundant items to prevent respondent fatigue and maintain data clarity. Additionally, correlations between survey items should be avoided to prevent skewing segmentation results.
- High-quality responses, free from biases and response styles such as extreme or midpoint tendencies, are essential. Ideally, responses should be binary or metric to facilitate straightforward analysis, particularly with distance-based techniques.
- Moreover, a sufficient sample size approximately 100 respondents per segmentation variable ensures robustness in identifying accurate market segments.

- These measures collectively enhance the reliability and effectiveness of market segmentation strategies.

3.4 Data from Internal Sources

- Organizations increasingly use internal data like scanner data and online purchases for segmentation due to its accuracy and automation. However, reliance solely on internal data risks bias towards existing customers, overlooking potential new customer behaviors.

3.5 Data from Experimental Studies

- Experimental data, derived from field or laboratory experiments such as advertising response tests or conjoint analyses, provides valuable insights for market segmentation. It assesses consumer preferences based on specific product attributes and their levels, offering criteria for segmenting markets effectively.

Step 5: Extracting Segments

5.1 Grouping Consumers

In market segmentation analysis, the choice of algorithm significantly shapes the resulting segmentation solution. Algorithms like k-means and single linkage hierarchical clustering impose different structures on consumer data. For instance, k-means tends to form round clusters, while single linkage can detect non-linear patterns like spirals. The suitability of an algorithm depends on data characteristics such as size, scale level of variables, and structural complexities. Therefore, exploring and comparing multiple methods is crucial to achieving a robust segmentation solution tailored to specific data nuances and segmentation objectives.

5.2 Distance-Based Methods

In market segmentation based on vacation activity patterns, understanding tourists' preferences using distance-based methods like Euclidean distance or cosine similarity is crucial. These methods quantify similarity or dissimilarity between individuals based on their activity profiles (e.g., BEACH, ACTION, CULTURE). Segmentation helps businesses tailor offerings to different tourist preferences, enhancing customer satisfaction and engagement during vacations.

5.2.1 Distance Measures

Euclidean Distance:

- Formula: d(x,y)=∑j=1p(xj−yj)2d(x, y) = \sqrt{\sum_{j=1}^{p} (x_j - y_j)^2}d(x,y)=∑j=1p(xj−yj)2
- Description: Measures the straight-line distance between two points in ppp-dimensional space.

Manhattan Distance (Absolute Distance):

- Formula: d(x,y)=∑j=1p∣xj−yj∣d(x, y) = \sum_{j=1}^{p} |x_j - y_j|d(x,y)=∑j=1p∣xj−yj∣

- Description: Computes distance as the sum of absolute differences along each dimension, similar to navigating city blocks in a grid.

  Asymmetric Binary Distance:

- Formula: d(x,y)=number of common 1snumber of 1s in x or yd(x, y) = \frac{\text{number of common 1s}}{\text{number of 1s in } x \text{ or } y}d(x,y)=number of 1s in x or ynumber of common 1s
- Description: Specifically for binary variables, measures the proportion of dimensions where both vectors have a value of 1.

Criteria for Distance Measures

- **Symmetry**: d(x,y)=d(y,x)d(x, y) = d(y, x)d(x,y)=d(y,x)
- **Identity**: d(x,y)=0 ⟺ x=yd(x, y) = 0 \iff x = yd(x,y)=0⟺x=y
- **Triangle Inequality**: d(x,z)≤d(x,y)+d(y,z)d(x, z) \leq d(x, y) + d(y, z)d(x,z)≤d(x,y)+d(y,z)

Practical Implementation in R

- Use `dist()` function in R to compute distances between rows of a data matrix, specifying `method = "euclidean"` or `method = "manhattan"`.

Application

- **Market Segmentation**: Helps identify groups of consumers with similar preferences or behaviors based on various metrics (e.g., vacation activities), enabling targeted marketing strategies.

5.2.2 Hierarchical Methods

- Hierarchical clustering methods are fundamental in data analysis for their intuitive approach to grouping data, which mirrors how humans naturally segment information. They are particularly effective in scenarios where understanding the hierarchical structure of data is crucial. These methods create a series of nested partitions, starting from one large cluster encompassing all observations to as many clusters as there are observations, each comprising a single entity. Divisive clustering begins with all observations in one group and recursively divides them into smaller clusters, while agglomerative clustering starts with each observation as a separate cluster and progressively merges the closest pairs. Both approaches yield a sequence of nested partitions, offering insights into the hierarchical organization of data. Key to their implementation are distance measures like Euclidean or Manhattan distances, and linkage methods such as single, complete, and average linkage, which determine how distances between groups of observations are calculated. In practice, these methods are implemented using the hclust() function in R, providing a deterministic algorithm that consistently generates the same sequence of partitions when applied to the same dataset, ensuring reliability and reproducibility in analysis.

- Underlying both divisive and agglomerative clustering is a measure of distance between groups of observations (segments). This measure is determined by specifying

A. (1) a distance measure d(x,y) between observations (consumers) x and y, and
B. (2) a linkage method. The linkage method generalises how, given a distance between pairs of observations, distances between groups of observations are obtained.

- Assuming two sets X and Y of observations (consumers), the following linkage methods are available in the standard R function hclust() for measuring the distance l(X,Y) between these two sets of observations: Single linkage: distance between the two closest observations of the two sets. l(X,Y) = min x∈X,y∈Y d(x,y) Complete linkage: distance between the two observations of the two sets that are farthest away from each other. l(X,Y) = max x∈X,y∈Y d(x,y) Average linkage: mean distance between observations of the two sets. l(X,Y) = 1 |X||Y| x∈Xy∈Y d(x,y), where |X| denotes the number of elements in X.
- Hierarchical clustering methods offer flexibility through different linkage methods and distance measures, allowing for varied exploratory analyses of data structures. The linkage methods—single, complete, and average—each approach cluster formation differently based on pairwise distances between observations.
- Single linkage, known for its "next neighbor" approach, merges clusters based on the closest observations, revealing non-linear and non-convex structures effectively. However, it can lead to chain effects where clusters connect simply because their nearest members are close.
- Complete linkage, in contrast, identifies clusters based on the maximum distance between any pair of observations from two sets, forming compact and well-separated clusters.
- Average linkage computes the mean distance between all pairs of observations from two sets, providing a balanced approach between single and complete linkage.
- Ward's method, utilizing squared Euclidean distances, minimizes variance within clusters by merging based on the minimal increase in total within-cluster variance, hence yielding compact and homogeneous clusters.
- Hierarchical clustering results are typically visualized as dendrograms, tree-like diagrams illustrating the merging process. Dendrograms show the hierarchy of clusters, with branch heights indicating distances between clusters. While dendrograms assist in determining the number of clusters, their utility can be limited when data lacks clear structure.
- Moreover, due to the potential for ties in distances between observations, different software packages may produce slightly varied dendrograms for the same dataset, impacting the visualization and interpretation of clustering solutions. Thus, understanding these nuances is crucial for accurate interpretation and application of hierarchical clustering in data analysis.
- This data set contains 563 respondents who state how often they take risks from the following six categories:
- 1. recreational risks: e.g., rock-climbing, scuba diving
- 2.health risks: e.g., smoking, poor diet, high alcohol consumption
- 3. career risks: e.g., quitting a job without another to go to
- 4. financial risks: e.g., gambling, risky investments
- safetyrisks:e.g.,speeding

- 6.socialrisks:e.g.,standing for election,publicly challenging a rule or decision.

5.2.3 Partitioning Methods

- Hierarchical clustering is best suited for small datasets with up to a few hundred observations due to the complexity of dendrogram interpretation and memory constraints with larger datasets. For larger datasets, partitioning clustering methods that compute distances only to cluster centers are more efficient and practical.

5.2.3.1 k-Means and k-Centroid Clustering

Initialization and Selection of Centroids:

1. The algorithm begins by specifying the desired number of segments kkk.
2. It randomly selects kkk observations (consumers) from the dataset to serve as initial cluster centroids C={c1,…,ck}C = \{c_1, \ldots, c_k\}C={c1,…,ck}.

Assigning Observations to Segments:

1. Each observation xix_ixi in the dataset is assigned to the closest cluster centroid cjc_jcj, forming initial segments S1,…,SkS_1, \ldots, S_kS1,…,Sk.
2. This assignment is based on minimizing the distance between xix_ixi and each centroid cjc_jcj.

Updating Cluster Centroids:

1. The centroids cjc_jcj are recalculated based on the current assignment of observations to segments.
2. For the squared Euclidean distance metric used in k-means, the centroids are updated as the mean of all observations assigned to each segment.

Iterative Refinement:

1. Steps 2 and 3 are repeated iteratively until convergence criteria are met.
2. Convergence occurs when the assignments of observations to segments and the centroids stabilize (i.e., minimal change between iterations) or when a maximum number of iterations is reached.

Convergence and Final Segmentation:

1. The iterative nature of the algorithm ensures that it converges to a solution, albeit not necessarily the global optimum.
2. Multiple runs with different initial centroid selections are recommended to find a more robust segmentation solution.

Challenges and Considerations:

1. Determining the optimal number of segments kkk remains a challenge and typically requires assessing stability across different kkk values.
2. The choice of distance metric (e.g., Euclidean, Manhattan) significantly influences the resulting segmentation solution, often more than the specific algorithm used.

5.2.3.2 Improved k-Means

- **Random Initialization Issues**: Randomly choosing initial centroids can lead to suboptimal clustering solutions because nearby centroids may not represent the entire dataset well.
- **Local Optima Problem**: Poor initialization increases the risk of k-means converging to local optima rather than the global optimum, affecting clustering quality.
- **Recommended Approach**: Use initialization methods that evenly distribute starting points across the data space to capture the dataset's true structure and improve clustering accuracy.
- **Research Finding**: Steinley and Brusco (2007) recommend randomly selecting multiple starting points and choosing the set that best represents the data to enhance k-means clustering performance.
- **Conclusion**: Effective initialization is crucial for achieving better clustering results with k-means by avoiding suboptimal solutions and enhancing overall accuracy.

5.2.3.3 Hard Competitive Learning

Hard competitive learning, also known as learning vector quantization, differs from k-means clustering in how segments are extracted. While both methods aim to minimize distances to segment centroids, hard competitive learning selects one consumer randomly to adjust its closest centroid, potentially yielding different segmentation outcomes even with identical starting points. It offers a distinct approach to finding segmentation solutions, potentially avoiding local optima encountered by k-means. Applications include segment-specific market basket analysis, demonstrated by Boztug and Reutterer (2008), and can be implemented in R using cclust(x, k, method = "hardcl") from the flexclust package.

5.2.3.4 Neural Gas and Topology Representing Networks

Neural gas, a variant of hard competitive learning, adjusts not only the nearest segment centroid towards a randomly selected consumer but also slightly adjusts the position of the second closest centroid, albeit to a lesser extent. This approach, introduced by Martinetz et al. (1993), has been applied in market segmentation studies, such as those by Dolnicar and Leisch (2010, 2014), and can be implemented in R using cclust(x, k, method = "neuralgas") from the flexclust package. A further development, topology representing networks (TRN), extends neural gas by organizing centroids based on their frequent co-adjustments with consumers, creating a virtual map that visualizes segment relationships. While these methods offer distinct approaches to clustering, they do not necessarily outperform k-means or hard competitive learning; rather, they provide alternative solutions for exploratory market segmentation analysis.

5.2.3.5 Self-Organising Maps

Self-organising maps (SOM), also known as Kohonen maps, position segment centroids on a grid (e.g., rectangular or hexagonal). Unlike other algorithms, SOM adjusts centroids and their neighboring centroids towards a randomly chosen consumer multiple times, gradually reducing adjustments until convergence. This structured approach aligns segment numbering with the grid but may increase distances between centroids and segment members due to grid-imposed location restrictions. SOM contrasts with k-means and other methods in segment numbering and has been applied in market segmentation studies, such as those discussed by Mazanec (1999) and Reutterer and Natter (2000). R packages like `kohonen` facilitate SOM implementation, offering visualisations of segment relationships on the grid.

5.2.3.6 Neural Networks

Auto-encoding neural networks for cluster analysis differ mathematically from traditional methods. They employ a single hidden layer perceptron where inputs are transformed into hidden layer nodes, and subsequently into output predictions. The network is trained to minimize the squared Euclidean distance between inputs and outputs, hence encoding data effectively. Once trained, the parameters connecting hidden and output layers serve as segment centroids, akin to traditional clustering. The hidden-to-input layer connections indicate segment membership probabilities, enabling fuzzy segmentation where consumers can belong to multiple segments. This contrasts with crisp segmentations of methods like k-means and hard competitive learning. R packages such as autoencoder facilitate implementation of auto-encoding neural networks for market segmentation (Dubossarsky and Tyshetskiy, 2015).

5.2.4 Hybrid Approaches

Hybrid segmentation approaches combine hierarchical and partitioning clustering algorithms to leverage their respective strengths. Hierarchical algorithms allow flexibility in segment number determination and provide visual insights via dendrograms. However, they require substantial memory and can be challenging with large datasets. In contrast, partitioning algorithms are memory efficient and suitable for large datasets but necessitate a predefined segment count and lack nested solutions. Hybrid methods start with a partitioning algorithm to handle large datasets, extracting numerous segments. Centroids and segment sizes are retained for hierarchical clustering, facilitating manageable data sizes for dendrogram analysis to determine the optimal number of segments.

5.2.4.1 Two-Step Clustering

IBM SPSS employs a method called two-step clustering, combining partitioning and hierarchical algorithms. Initially, a partitioning procedure like k-means extracts a large number of clusters. Then, centroids and segment sizes from this step are retained for hierarchical clustering to determine the optimal number of segments. This hybrid approach has been applied across various domains, such as mobile phone usage types and electric vehicle adoption patterns. In R, similar steps can be implemented manually using functions like `stepcclust()` for partitioning and `hclust()` for hierarchical clustering, offering flexibility beyond commercial software limitations.

5.2.4.2 Bagged Clustering

· **Methodology of Bagged Clustering**:

- Combines hierarchical and partitioning clustering algorithms with bootstrapping (Efron and Tibshirani, 1993).
- Bootstrapping involves randomly drawing data with replacement to create multiple bootstrap samples, reducing dependency on specific data points.
- Begins with partitioning clustering on bootstrap samples, followed by hierarchical clustering on resultant cluster centroids.
- Hierarchical clustering helps determine optimal segment number from dendrogram structure.

· **Advantages and Applicability**:

- Effective for identifying niche markets and avoiding local optima.
- Suitable for large datasets by discarding original data post bootstrap, using centroids.
- Provides deterministic segmentation using hierarchical clustering.

· **Applications in Tourism**:

- Successfully applied in tourism to segment winter vacation activities data.
- Example dataset: Austrian winter tourist activities survey (1997/1998), 27 activities considered.
- Identified segments based on diverse activities, including skiing, spa visits, and cultural outings.

· **Steps of Bagged Clustering**:

- Create b bootstrap samples from dataset X.
- Apply partitioning method (e.g., k-means) to each sample, producing b × k cluster centroids.
- Use these centroids as new data, perform hierarchical clustering.

- Final segmentation determined by dendrogram cut, assigning consumers to nearest segment.

· **Visualization and Analysis**:

- Utilizes dendrograms and bar charts to visualize segment structures and sizes.
- Boxplots illustrate variability across segments for various vacation activities.
- Insights gained into segment characteristics, like preference for skiing vs. spa activities.

· **Statistical Considerations**:

Built-in uncertainty analysis due to ensemble nature (multiple bootstrap samples).

- Results sensitive to random number generator changes over time but qualitatively consistent.

5.3 Model-Based Methods

· **Introduction to Model-Based Methods**:

- · Model-based methods are an alternative to distance-based clustering for market segmentation.
- Proposed by Wedel and Kamakura (2000), they involve mixture methodologies, which are gaining popularity.
- Offer a different approach to segment extraction based on statistical models rather than similarity metrics.

· **Principles of Model-Based Segment Extraction**:

- Unlike distance-based methods, model-based methods assume:
    - Each segment has a specific size (πh).
    - Consumers in the same segment share common characteristics specific to that segment (θh).
- Employs finite mixture models where the model is a combination of segment-specific models.

· **Estimation Techniques**:

- Parameters (segment sizes π and segment-specific characteristics θ) are estimated using statistical methods like maximum likelihood estimation.
- The EM algorithm is commonly used due to its iterative nature and ability to handle missing data (segment memberships).

· **Determining Number of Segments**:

- Selecting the optimal number of segments (k) is crucial but challenging.
- Information criteria such as AIC, BIC, and ICL help in choosing k by balancing model fit and complexity.
- These criteria penalize complex models to avoid overfitting.

· **Segment Assignment**:

- Consumers are assigned to segments based on the highest probability of segment membership (Eq. 7.2).
- Mean entropy measures the clarity of segment assignments, where lower entropy indicates clearer assignments.

· **Advantages and Extensions**:

- Model-based methods can capture complex segment characteristics better than distance-based methods.
- Extensions include varying segment-specific models beyond θ, allowing for more nuanced segmentation.

· **Literature and Complexity**:

- Extensive literature on finite mixture models supports their flexibility and applicability in diverse segmentation scenarios.
- Terminology includes mixture components (segments), prior probabilities (segment sizes), and posterior probabilities (segment membership probabilities).

· **Statistical Frameworks**:

- Bayesian methods also applicable, using Markov chain Monte Carlo techniques for parameter estimation.

7.3.1 Finite Mixtures of Distributions

The simplest case of model-based clustering has no independent variables x, and simply fits a distribution to y. To compare this with distance-based methods, finite mixtures of distributions basically use the same segmentation variables: a number of pieces of information about consumers, such as the activities they engage in when on vacation. No additional information about these consumers, such as total travel expenditures, is simultaneously included in the model. The finite mixture model reduces to k h=1 πhf(y|θh), πh ≥ 0, k h=1 πh =1. (7.6) TheformulaearethesameasinEq.7.1, the only difference is that there is no x. The statistical distribution function f()depends on the measurement level or scale of the segmentation variables y

5.3.1.1 Normal Distributions

- In summary, model-based methods such as finite mixture models provide a robust alternative to traditional distance-based clustering approaches for market segmentation. These methods assume distinct segment characteristics and sizes within empirical data and use statistical models to identify these segments effectively.
- Finite mixture models, particularly those employing multivariate normal distributions, excel in capturing intricate relationships among segmentation variables. They can model dependencies among variables, such as correlations in consumer spending across different product categories or physical measurements like height and limb proportions. This capability is crucial for understanding complex market behaviors where variables are interconnected, like in competitive pricing or physical attribute correlations.
- The process of fitting a finite mixture model involves estimating parameters like segment means and covariance matrices. This estimation is typically done using methods such as maximum likelihood or Bayesian approaches. The selection of the number of segments (clusters) is pivotal and often guided by information criteria such as AIC or BIC, which balance model complexity with how well the model fits the data.
- When applying finite mixture models to practical datasets—such as consumer behavior or vacation motives—the choice of covariance model (e.g., spherical, ellipsoidal) influences both the interpretability and computational efficiency of the segmentation model. Visual tools like uncertainty plots and classification plots are valuable for assessing the model's performance, identifying ambiguous segment assignments, and detecting artificially created segments due to model constraints.
- In conclusion, while finite mixture models require careful parameter tuning and validation, they provide a powerful framework for uncovering hidden structures within data. They are particularly well-suited for scenarios where segment characteristics are distinct yet interconnected across multiple dimensions. This approach represents a sophisticated toolset for modern market segmentation analysis, capable of handling the intricate complexities present in diverse datasets effectively.

5.3.1.2 BinaryDistributions

Finite mixture models, such as those applied to binary data in market segmentation, are powerful tools for identifying distinct consumer segments based on binary preferences or behaviors. These models assume that segments have different probabilities of engaging in specific activities, capturing complex relationships that traditional clustering methods may miss. By fitting these models using algorithms like EM, we can uncover latent segments and their characteristic behaviors, informing targeted marketing strategies tailored to different consumer preferences.

5.3.2 FiniteMixturesofRegressions

Finite mixtures of regressions (FMR) represent a sophisticated statistical approach for segmenting data where the relationship between a dependent variable (like willingness to pay) and independent variables (predictors such as number of rides in a theme park) varies across different segments or clusters. Here's a concise overview:

Concept: FMR assume that data arises from multiple subpopulations, each characterized by its own regression model. This allows for heterogeneity in regression relationships across segments.

Application: Useful in market segmentation where different consumer groups may exhibit distinct preferences or behaviors that influence their response to predictors.

Example: In the theme park scenario, FMR identified two segments based on willingness to pay: one segment had a linear relationship between pay and rides, while another had a quadratic relationship, illustrating different valuation behaviors.

Modeling: Implemented using packages like flexmix in R, FMR fits multiple regression models simultaneously and assigns each observation to the segment where it fits best.

Advantages: Provides more nuanced insights than traditional clustering methods by capturing varying regression relationships within data segments.

Challenges: Computational complexity and the need for careful model selection (e.g., determining the optimal number of segments).

Interpretation: Results in segment-specific regression coefficients that describe how predictors affect the dependent variable differently across segments.

5.3.3 Extensions and Variations

Finite mixture models are powerful tools in statistical modeling, offering a flexible approach to segmenting data by accommodating various data types and characteristics. They can handle metric data with mixtures of normal distributions, binary data with mixtures of binary distributions, and categorical data with mixtures of multinomial distributions or multinomial logit models. For ordinal variables susceptible to response styles, mixture models effectively disentangle response biases from genuine segment differences, enhancing segmentation accuracy. They integrate well with conjoint analysis to account for diverse consumer preferences across segments. Moreover, these models reconcile the debate between continuous and discrete segmentation by recognizing distinct segments while allowing for variability within each segment. Extensions like mixture of mixed-effects models further refine segmentation by incorporating both segment-specific effects and individual-level variations. For dynamic behaviors in time-series data, mixture models using Markov chains or dynamic latent change models track evolving consumer behaviors over time, crucial for understanding brand choices and purchasing decisions. Additionally, by including descriptor variables, such as demographics, in modeling segment sizes, mixture models provide a comprehensive framework to capture how segments differ compositionally in response to external factors.

5.4 Algorithms with Integrated Variable Selection

Biclustering Algorithms

Biclustering algorithms are designed to simultaneously cluster consumers and variables. They are particularly useful when dealing with binary data, where the goal is to identify segments of consumers who share common traits across a subset of variables. Unlike traditional clustering methods, biclustering considers both rows (consumers) and columns (variables) together, aiming to find cohesive groups where all members exhibit specific patterns of binary responses.

Key points:

- **Definition and Purpose**: Biclustering identifies biclusters—groups of consumers who share 1s across a subset of binary variables.
- **Applications**: Originally popularized in genetic and proteomic data analysis, where identifying groups sharing genetic or protein expression patterns is crucial.
- **Algorithm**: Various algorithms exist with different criteria for defining biclusters. For instance, the Bimax algorithm, proposed by Prelic et al. (2006), is known for efficiently identifying the largest rectangles of 1s in binary matrices.
- **Advantages**: Suitable for identifying niche markets and patterns in high-dimensional data without prior data transformation, which could bias segmentation results.

Variable Selection Procedure for Clustering Binary Data (VSBD)

The Variable Selection Procedure for Clustering Binary Data (VSBD), proposed by Brusco (2004), aims to optimize segmentation by selecting relevant variables. It's particularly useful when dealing with high-dimensional binary datasets, where many variables may not contribute significantly to the clustering solution.

Key points:

- **Objective**: Identify a subset of variables that optimally differentiate between clusters.
- **Method**: Based on the k-means clustering algorithm, it iteratively adds variables that minimize the within-cluster sum-of-squares criterion until a threshold increase is reached.
- **Advantages**: Reduces computational complexity and improves interpretability by focusing on the most relevant variables.
- **Implementation**: Available in software packages like flexclust in R, where users can specify parameters such as the number of clusters and thresholds for variable inclusion.

Factor-Cluster Analysis

Factor-cluster analysis combines factor analysis with cluster analysis in a two-step approach. First, segmentation variables are reduced to factors through factor analysis. Then, these factor scores are used to perform cluster analysis to identify market segments.

Key points:

- **Purpose**: Reduces the dimensionality of segmentation variables to improve computational efficiency and interpretability.
- **Drawbacks**: Results in loss of information because it discards original variable details, potentially biasing segmentation outcomes.
- **Suitability**: Recommended for cases where the number of segmentation variables is prohibitively large relative to sample size, though it alters data fundamentally.

These methods cater to different needs in market segmentation, from handling high-dimensional binary data efficiently to reducing the complexity of variable sets for clearer segmentation insights. Each approach offers distinct advantages depending on the nature and objectives of the segmentation analysis.

5.5.1 Cluster Indices

Internal Cluster Indices

Internal cluster indices are used to evaluate a single segmentation solution without relying on external information.

Sum of Within-Cluster Distances (Wk):

1. Measures the compactness of clusters by calculating the sum of distances between each segment member and their segment representative (centroid).
2. Lower Wk indicates more similar members within the same segment.
3. Used to assess how well-separated segments are in terms of distance.

Ball-Hall Index (Wk/k):

1. A variation of Wk that normalizes the sum of within-cluster distances by dividing it by the number of segments (k).
2. Corrects for the tendency of Wk to decrease monotonically as k increases.

Between-Cluster Distances (Bk):

1. Measures the dissimilarity between clusters using weighted distances between centroids.
2. Evaluates how different clusters are from each other.

Combination Indices:

1. Combine Wk and Bk to assess both compactness and separation of clusters.
2. Useful for determining the optimal number of segments based on internal measures.

Calinski-Harabasz Index (CHk):

1. Specifically recommended for identifying the optimal number of segments.
2. Ratio of between-cluster distance to within-cluster distance, adjusted for sample size.
3. Higher CHk values indicate better-defined clusters.

Scree Plot:

1. Graphical representation of Wk across different numbers of segments (k).
2. Helps identify an "elbow" where the rate of decrease in Wk changes, suggesting an optimal number of segments.

External Cluster Indices

External cluster indices require additional external information to evaluate segmentation solutions.

Jaccard Index:

1. Compares similarity between two segmentation solutions based on the proportion of pairs of observations assigned to the same segment.
2. Value ranges from 0 (completely different) to 1 (identical).
3. Useful when true segment structure (known segmentation) is available.

Rand Index:

1. Similar to the Jaccard index but considers all possible pairs of observations.
2. Also ranges from 0 to 1, where 1 indicates perfect agreement.

Adjusted Rand Index:

1. Corrects for the effect of chance in the Rand index by adjusting for segment sizes.
2. Takes into account the expected agreement by chance given segment sizes.
3. Provides a more reliable measure of similarity between segmentation solutions, especially when segment sizes differ.

Usage and Interpretation

- **Applications**: Used to guide decisions on the number of segments to extract in market segmentation analysis.
- **Tools**: Various packages in R (e.g., fpc, clusterSim, NbClust, flexclust) provide functions to compute these indices.
- **Challenges**: Internal indices may not always provide clear guidance, especially in consumer data lacking natural segment structures, where external indices and stability analysis become crucial.

5.5.2 Gorge Plots

In market segmentation analysis, gorge plots visualize similarity values between consumers and segment representatives. They help assess how well-defined segments are, with high values indicating proximity to the segment centroid and low values indicating distance. Gorge plots ideally show a peak on both ends, resembling a gorge, suggesting well-separated segments.

5.5.3 Global Stability Analysis

resampling methods, such as bootstrapping, are crucial for assessing the stability and reliability of market segmentation solutions derived from consumer data:

Purpose: Evaluate segmentation robustness amidst data variability and algorithm sensitivity.

Procedure: Generate multiple datasets via bootstrap sampling, apply segmentation algorithms, and compare solutions using metrics like Adjusted Rand Index.

Insights: Identify natural segments (high stability), reproducible structures (moderate stability), or constructive segments (low stability).

Applications: Facilitates informed segmentation decisions by aligning data-driven insights with strategic business goals.

Segment Level Stability Analysis

Importance of Segment Level Stability (SLSW):

1. **Purpose**: To assess the stability of individual segments within a segmentation solution rather than the entire solution.
2. **Rationale**: Ensures that even if the overall segmentation solution is not highly stable, specific segments of interest (e.g., niche markets) might still be stable and worth considering.
3. **Method**: Uses bootstrap sampling to create multiple segmentation solutions and measures agreement between segments across these solutions.

Procedure for SLSW:

1. **Steps**:
    1. Compute a partition (segmentation solution) using a clustering algorithm.
    2. Generate bootstrap samples from the original data.
    3. Cluster each bootstrap sample into segments.

4. Measure agreement (Jaccard index) between segments across bootstrap samples.
5. Assess segment stability using boxplots of Jaccard coefficients across bootstrap samples.

Example with Artificial Mobile Phone Data:

1. **Data**: Contains artificially generated segments (three distinct segments).
2. **Analysis**:
    1. Segmentation into three and six segments.
    2. slswFlexclust function in R used to calculate SLSW for each segmentation solution.
    3. Visual representation (boxplots) shows segment stability (Fig. 7.42).
    4. Interpretation: Higher stability (Jaccard index near 1) indicates robust segments; lower stability indicates unstable or artificial segments.

Real-World Example: Australian Travel Motives:

1. **Data**: Survey data with 20 travel motives of Australian residents.
2. **Method**: Neural gas algorithm used for clustering.
3. **Analysis**:
    1. Segmentation into multiple solutions (three to eight segments).
    2. slswFlexclust used to assess SLSW for each solution.
    3. Segment stability visualized (Fig. 7.43) and interpreted.
    4. Conclusion drawn based on stability metrics: Some segments (e.g., segment 6) show high stability across solutions, indicating natural segments; others (e.g., segment 3) show instability, likely artifacts of the analysis.

Segment Level Stability Across Solutions (SLSA)

Definition and Purpose:

1. **SLSA**: Evaluates how consistently segments appear across different segmentation solutions with varying numbers of segments.
2. **Indicator**: High SLSA values suggest natural segments present in the data, rather than artifacts of the clustering process.

Procedure for SLSA:

1. **Steps**:
    1. Calculate SLSA for a series of partitions (segmentation solutions) with different numbers of segments.
    2. Assess how segments transition between different solutions.
    3. Visualize with plots (e.g., Fig. 7.44, Fig. 7.45).

4. Use entropy to quantify stability numerically.

Application Examples:

1. **Artificial Mobile Phone Data**: Shows segment transitions across three to eight segments (Fig. 7.44).
2. **Australian Travel Motives**: Demonstrates stability patterns across different solutions (Fig. 7.45).
3. **Interpretation**: Segments with high stability across solutions are more likely to represent natural market segments.

B Pavan Srinivasarao

Step1

Step2

Step3

Step6

Step9

Step 1 : Deciding (not) to Segment

1)Market segmentation is a key marketing strategy applied many organizations , it isnot the best decision there will be many implications in it it is either good or bad 2)the market segmentation includes the the creation of the new products,the modification of existing products,changes in pricing and distribution channels used tosell the product, as well as all communications with the market

3)this makes changes in internal structure of the organization which may be adjusted 4)the major implications of such a long-term organizational commitment, the decision toinvestigate the potential of a market segmentation strategy must be made at the highestexecutive level, and must be systematically and continuously communicated and reinforced at all organizational levels and across all organizational units.

3.2 Implementation Barriers:

As many plans were before executing there will be some barriers and in market segmentation there are some :

1) The first group of barriers relates to senior management. Lack of leadership,pro-active championing, commitment and involvement in the market segmentation
2) Senior management can also prevent market segmentation to be successfully implemented by not making enough resources available, either for the initial market segmentation analysis itself, or for the long-term implementation of a market segmentation strategy
3) Organizational culture barriers such as lack of market orientation, resistance to change, poor communication, and short-term thinking can hinder the successful implementation of market segmentation, as identified by Dibb and Simkin (2008),with Croft (1994) developing a questionnaire to assess these cultural

   impediments.
4) The lack of a qualified data manager and analyst in the organization can also represent major stumbling blocks
5) Another obstacle may be objective restrictions faced by the organisation, including lack of financial resources, or the inability to make the structural changes required.
6) Most of these barriers can be identified from the outset of a market segmentation study, and then proactively removed. If barriers cannot be removed, the option of abandoning the attempt of exploring market segmentation as a potential future strategy should be seriously considered.

7) If the company is not a marketing company, implementation of the marketing segmentation is not easy

Step 2: Specifying the Ideal Target Segment:

Segment Evaluation Criteria

1) The third layer of market segmentation analysis depends on the user input user input cannot be limited to either a briefing at the start of the process, or the development of a marketing mix at the end

2) Th step 2 can be done in two steps :
   a) One set of evaluation criteria can be referred to as knock-out criteria. These criteria are the essential, non-negotiable features of segments that the organisation would consider targeting
   b) The second set of evaluation criteria can be referred to as attractiveness criteria. These criteria are used to evaluate the relative attractiveness of the remaining market segments – those in compliance with the knock-out criteria

Knock-Out Criteria and Attractiveness Criteria

1) Knock-out criteria are used to determine if market segments resulting from the market segmentation analysis qualify to be assessed using segment attractiveness criteria.
2) Knock-out criteria must be understood by senior management, the segmentationteam, and the advisory committee
3) To implement the knock-out-criteria they require the segment must be homogeneous,distinct,large enough,matching,identfiable ,reachable
4) Attractiveness criteria are not binary in nature. Segments are not assessed as either complying or not complying with attractiveness criteria. Rather, each market segment israted

1) To achieve this, a large number of possible criteria has to be investigated before agreement is reached on which criteria are most important for the organisation. McDonald and Dunbar (2012) recommend to use no more than six factors as the basisfor calculating these criteria

2) At the end of this step, the market segmentation team should have a list of approximately six segment attractiveness criteria. Each of these criteria should have aweight attached to it to indicate how important it is to the organisation compared to theother criteria.

STEP 3 : COLLECTING DATA

Segmentation Variables:

1) The segmentation is nothing but the one consumer than the character of that represent the descriptor variables , the descriptor variables gives the more informationabout the consumer or segment

2)Describing segments is critical to being able to develop an effective marketing mix targeting the segment. Typical descriptor variables include socio-demographics, butalso information about media behaviour, allowing marketers to reach their target segment with communication messages

3)The socio-demographic variables, gender, age, and the number of vacations undertaken per annum serve as descriptor variables

4)The correct description, in turn, makes it possible to develop a customised product, determine the most appropriate pricing strategy, select the best distribution channel, and the most effective communication channel for advertising and promotion.

Segmentation Criteria:

1) Long before segments are extracted, and long before data for segment extraction is collected, the organisation must make an important decision: it must choose which segmentation criterion to use

2) The most common segmentation criteria are geographic, sociodemographic, psychographic and behavioural.

A) Geographic Segmentation

1) The Geographic Segmentation includes classify the customers based on the regionaland promoting the things they want

2) it is very easy to identify the customers they want then easily applicable to that areabut they will be another loss the persons who are living in that area having of chance not liking their traditional areas then in that situation the geographic segmentation willnot work

B)Socio-Demographic Segmentation

1) Typical socio-demographic segmentation criteria include age, gender, income and education. Socio-demographic segments can be very useful in some industries.

2) In many instances, the socio-demographic criterion is not the cause for product preferences, thus not providing sufficient market insight for optimal segmentation decisions.

3) Haley (1985) estimates that demographics explain about 5% of the variance in consumer behaviour

4) values, tastes and preferences are more useful because they are more influential interms of consumers' buying decisions. 44 5 Step 3: Collecting Dat

C)Psychographic Segmentation

1) When people are grouped according to psychological criteria, such as their beliefs, interests, preferences, aspirations, or benefits sought when purchasing a product, theterm psychographic segmentation is used.
2) Lifestyle segmentation is another popular psychographic segmentation approach it isbased on people’s activities, opinions and interests
3) Psychographic criteria are, by nature, more complex than geographic or sociodemographic criteria because it is difficult to find a single characteristic of a personthat will provide insight into the psychographic dimension of interest.
4) The disadvantage of the psychographic approach is the increased complexity of determining segment memberships for consumers. Also, the power of the psychographic approach depends heavily on the reliability and validity of the empiricalmeasures used to capture the psychographic dimensions of interest

D)Behavioural Segmentation

1) The key advantage of behavioural approaches is that – if based on actual behaviourrather than stated behaviour or stated intended behaviour – the very behaviour of interest is used as the basis of segment extraction

2) But behavioural data is not always readily available, especially if the aim is to includein the segmentation analysis potential customers who have not previously purchased the product

Data from Survey Studies

The survey data were very cheap to collect but purely depends upon the Survery it willbe more negative impact on this

1) the survey reports depends on the following steps
   a) Choice of variables
   b) Response styles
   c) Response options
   d) Sample size of the data collect

Data from the internal sources :

1) Increasingly organisations have access to substantial amounts of internal data thatcan be harvested for the purpose of market segmentation analysis
2) Another advantage is that such data are usually automatically generated and – if organisations are capable of storing data in a format that makes them easy to access –no extra effort is required to collect data.
3) The danger of using internal data is that it may be systematically biased byover-representing existing customers

Data from Experimental Studies

1) Another possible source of data that can form the basis of market segmentation analysis is experimental data. Experimental data can result from field or laboratoryexperiments.
2) Experimental data can also result from choice experiments or conjoint analyses
3) Conjoint studies and choice experiments result in information about the extent to which each attribute and attribute level affects choice. This information can also be usedas a segmentation criterion

STEP 6 PROFILING SEGMENTS

Identifying Key Characteristics of Market Segments:

1) The process of profiling market segments involves identifying and characterizing thekey characteristics of each segment derived from data-driven segmentation.
They were some steps to follow in the profiling segments :

A) Identification of Characteristics: Profiling involves identifying the defining characteristics of each market segment with respect to the segmentation variables used**B)Comparison Across Segments**: It's not just about describing each segment individually but also comparing them to understand how they differ. This comparisonhelps in identifying unique traits that differentiate one segment from another.

C) Alternative Segmentation Solutions: Sometimes, multiple segmentation solutionsneed to be explored to find the most meaningful and actionable segments. This is particularly important when the data doesn't naturally form clear segments.
D) Interpretation Challenges: Data-driven segmentation results can be complex to interpret, as highlighted by studies showing managerial difficulties in understanding andusing segmentation insights effectively. Graphical statistical approaches can aid in presenting segmentation profiles more clearly and intuitively.

E) Graphical Statistical Approaches: These approaches include techniques like cluster analysis dendrograms, perceptual maps, and other visualizations that summarize complex segmentation results. They make the profiling process less tediousand reduce the risk of misinterpretation by presenting the data in a more accessible format.

Traditional Approaches to Profiling Market Segments:

1) The data is to collected from the different data warehouses then that data normally analyzeand give the descriptive data but giving of the descriptive data does not clarity of the data analysis to the person

2) Sometimes – to deal with the size of this task – information is provided aboutthe statistical significance of the difference between segments for each of the segmentation variables.

3) This approach, however, is not statistically correct Segment membership is directly derived from the segmentation variables, and segments are created in a way that makes them maximally different, thus not allowing the use of standard statistical tests toassess the significance of differences.

Segment Profiling with Visualisations:

1) Visualisations are useful in the data-driven market segmentation process to inspect, for each segmentation solution, one or more segments in detail.
2) Statistical graphs facilitate the interpretation of segment profiles. They also make it easier to assess the usefulness of a market segmentation solution.
3) The process of segmenting data always leads to a large number of alternative solutions. Selecting one of the possible solutions is a critical decision.
4) Visualisations of solutions assist the data analyst and user with this task

Identifying Defining Characteristics of Market Segments

1) **Segment Profile Plot**: Visual representation of segmentation variables showing how segments differ from the overall sample. It's a more intuitive and faster method for interpreting data compared to tables.

2) **Ordering Variables**: Variables can be ordered by their meaningful order in the dataset or rearranged to improve visualization. Clustering of columns can be achieved through hierarchical clustering.
3) **Hierarchical Clustering**: Variables are clustered using Ward's method to identify similar answer patterns. This clustering helps in reordering variables to enhance visualization.
4) **Marker Variables**: In the segment profile plot, marker variables (characteristic variables for a segment) are highlighted in color, while other variables are greyed out. Marker variables are defined based on their absolute and relative differences from the overall mean.
5) **Interpretation and Thresholds**: Empirically determined thresholds (0.25 and 50% relative difference) help in identifying marker variables. Different thresholds may be needed for non-binary variables.
6) **Comparison with Tables**: The segment profile plot is easier and faster to interpret than tables. It provides the same information but in a visually accessible format.
7) **Eye Tracking Study**: A study showed that segment profile plots require less cognitive effort to interpret compared to traditional or improved tables. This makes them valuable for managers making strategic decisions

Assessing Segment Separation:

1) Segment separation plots are very simple if the number of segmentation variables is low, but become complex as the number of segmentation variables increases. But evenin such complex situations, segment separation plots offer data analysts and users a quick overview of the data situation, and the segmentation solution.

A) Multiple Projections: Segment separation can be visualized through various projections of the data. Each projection may reveal different aspects of segment overlap or separation. It's important to examine multiple projections to get a comprehensive understanding of segment distinctiveness.

B) Single Projection Insights: While a single projection can indicate that certain segments are well-separated, it should not be taken as definitive proof of overall segment separation. Well-separated segments in one projection suggest distinct differences in the specific attributes visualized, but other projections might show overlap.

C) Projection Overlap: Overlap in a specific projection does not necessarily mean that segments are indistinguishable in all projections. Segments 1 and 5 might overlap in one projection but could be well-separated in another based on different variables or dimensions.

D) Distinct Segments: When segments are well-separated in a projection, it indicates clear differences in the specific characteristics or behaviors being analyzed. For instance, segments 6 and 3 being well-separated in a travel motives projection suggest they represent tourists with distinctly different travel preferences.

E) Visual and Analytical Assessment: Visual tools like segment separation plots helpin assessing how distinct the segments are. However, it is also important to use analytical methods such as statistical tests or clustering validation indices to quantitatively assess the degree of separation and validate the robustness of the segmentation.

STEP 9: Customising the Marketing Mix

Implications for marketing for mixing decisions :

1) Marketing means nothing experiment of all market ingredients to get more profit and long durable product previously traditions includes 12 ingredients they are product planning, packaging, physical handling, distribution channels, pricing, personal selling, branding, display, advertising, promotions, servicing, fact finding and analysis
2) On going years it were decreased to 4 ingredients of market they are Product, Price, Promotion and Place
3) The selection of one or more specific target segments may require the design of new, or the modification or re-branding of existing products (Product), changes to prices or discount structures (Price), the selection of suitable distribution channels (Place), and the development of new communication messages and promotion strategies that are attractive to the target segment (Promotion).
4) If the market segmentation analysis is conducted for the purpose of informing distribution decisions, store loyalty, store patronage, and benefits sought when selectinga store may represent valuable segmentation variables

Product:

The key decisions in developing the product dimension of the marketing mix involve aligning the product with customer needs, often through modifying existing products rather than creating entirely new ones. Other product-related decisions include naming, packaging, offering warranties, and providing after-sales support.

For targeting segment 3, potential product strategies include:

1. Developing a new product such as "MUSEUMS, MONUMENTS & MUCH, MUCH MORE" accompanied by an activities pass, facilitating easy discovery of relevant activities during vacation planning.
2. Enhancing gardens at the destination to become a primary attraction for this segment.

These strategies illustrate how product design or modification can be tailored to meet the specific preferences and behaviors of target market segments identified through biclustering analysis.

PRICE :

Typical decisions an organisation needs to make when developing the price dimension of the marketing mix include setting the price for a product, and deciding on discounts tobe offered.

Place :

1)The key decision relating to the place dimension of the marketing mix is how to distribute the product to the customers.
2) This includes answering questions such as: should the product be made available for purchase online or offline only or both; should the manufacturer sell directly to customers; or should a wholesaler or a retailer or both be used

Promotion:

1)Typical promotion decisions that need to be made when designing a marketing mix include: developing an advertising message that will resonate with the target market, and identifying the most effective way of communicating this message.
2) Other tools in the promotion category of the marketing mix include public relations, personal selling, and sponsorship.

Himesh Ray .

Step1

Step2

Step4

Step8

Summary and Important Points on Market Segmentation

Step 1: Deciding (not) to Segment:

1. Long-term Commitment

- Market segmentation requires a long-term commitment and substantial changes in the organization.
- It involves costs related to research, surveys, designing packages, and advertisements.
- Organizations should only pursue segmentation if the expected increase in sales justifies the costs.

2. Organizational Changes

- New products, modification of existing products, changes in pricing, distribution channels, and communication strategies may be needed.
- These changes can impact the internal structure of the organization, requiring adjustments to target different market segments.
- Strategic business units focused on market segments may be necessary.

3. Executive-Level Decision

- The decision to investigate market segmentation must be made at the highest executive level.
- It must be systematically communicated and reinforced across all organizational levels.

4. Implementation Barriers

- **Senior Management**: Lack of leadership, commitment, and resources can undermine segmentation efforts.
- **Organizational Culture**: Resistance to change, lack of market orientation, poor communication, and short-term thinking are significant barriers.
- **Lack of Training**: Inadequate understanding of market segmentation among senior management and the segmentation team can lead to failure.
- **Absence of Marketing Expertise**: Lack of a formal marketing function or qualified marketing expert can hinder implementation.
- **Financial Constraints**: Limited financial resources and inability to make necessary structural changes can impede segmentation efforts.
- **Process-Related Barriers**: Unclear objectives, lack of planning, and inadequate structured processes can obstruct successful implementation.
- **Operational Challenges**: Difficulty in understanding and using management science techniques can prevent successful adoption.

5. Proactive Removal of Barriers

- Most barriers can be identified and removed proactively.
- If barriers cannot be removed, the option of abandoning the segmentation strategy should be considered.

6. Checklist for Deciding to Segment

- **Culture Assessment**: Ensure the organization is market-oriented, willing to change, open to new ideas, and has good communication across units.
- **Resource Assessment**: Confirm the organization can make significant structural changes and has sufficient financial resources.
- **Senior Management Commitment**: Secure visible and active commitment, financial support, and understanding of the segmentation concept from senior management.
- **Team Formation**: Assemble a team with marketing, data, and data analysis experts. Set up an advisory committee representing all organizational units.
- **Objective Clarity**: Ensure the objectives of the market segmentation analysis are clear and responsibilities are assigned.
- **Structured Process**: Develop a structured process to follow and ensure there is enough time to conduct the analysis without pressure.

Summary of Key Points

- Market segmentation is a long-term strategic commitment requiring substantial changes and resources.
- Successful implementation depends on strong leadership, a supportive organizational culture, adequate training, and sufficient resources.
- Barriers to implementation can be proactively addressed, but if insurmountable, segmentation should be reconsidered.
- A clear and structured approach with committed senior management and a competent team is essential for successful market segmentation.

Step 2: Specifying the Ideal Target

Segment

Segment Evaluation Criteria

1. User Involvement

- User input is crucial at multiple stages of market segmentation analysis.

- Continuous involvement from the organization is needed, not just at the beginning or end of the process.

2. Conceptual Contribution

- After committing to segmentation in Step 1, the organization must significantly contribute to Step 2.
- This involves setting up two sets of evaluation criteria: **knock-out criteria** and **attractiveness criteria**.

3. Knock-Out Criteria

- Essential and non-negotiable features that a target segment must have.
- These criteria eliminate segments that do not meet the minimum requirements.

4. Attractiveness Criteria

- Used to evaluate the relative attractiveness of the remaining segments.
- These criteria are negotiable and varied, reflecting what makes a segment appealing to the organization.
- The segmentation team selects and prioritizes these criteria based on the organization's needs.

5.Differentiating Criteria

- Knock-out criteria are essential and non-negotiable.
- Attractiveness criteria are flexible and tailored to evaluate the potential of each segment.

6. Application of Criteria

- Knock-out criteria are used to filter out unsuitable segments automatically.
- Attractiveness criteria are applied to assess and rank the remaining segments based on their relative appeal to the organization.

Step 3 Collecting Data

Segmentation Variables

The term **segmentation variable** refers to the variable in the empirical data used in commonsense segmentation to split the sample into market segments. In commonsense segmentation, the segmentation variable is typically one single characteristic of the consumers in the sample.

All the other personal characteristics available in the data – in this case: age, the number of vacations taken, and information about five benefits people seek or do not seek when they go on vacation – serve as so-called **descriptor variables**. They are used to describe the segments in detail. Describing segments is critical to being able to develop an effective marketing mix targeting the segment. Typical descriptor variables include socio-demographics and media behavior information, allowing marketers to reach their target segment with communication messages.

Difference Between Commonsense and Data-Driven Market Segmentation

The difference between commonsense and data-driven market segmentation is that data-driven market segmentation is based not on one, but on multiple segmentation variables. These segmentation variables serve as the starting point for identifying naturally existing or artificially creating market segments useful to the organization.

Importance of Data Quality

These two examples illustrate how critical the quality of empirical data is for developing a valid segmentation solution. When commonsense segments are extracted, even if the nature of the segments is known in advance, data quality is critical to both:

1. Assigning each person in the sample to the correct market segment.
2. Being able to correctly describe the segments.

The correct description makes it possible to develop a customized product, determine the most appropriate pricing strategy, select the best distribution channel, and the most effective communication channel for advertising and promotion.

Data Sources for Segmentation Studies

Empirical data for segmentation studies can come from a range of sources: surveys, observations such as scanner data where purchases are recorded and linked to an individual customer's long-term purchase history via loyalty programs, or experimental studies. Optimally, data used in segmentation studies should reflect consumer behavior.

Segmentation Criteria

The term **segmentation criterion** is used here in a broader sense than the term segmentation variable. The term segmentation variable refers to one measured value, such as one item in a survey, or one observed expenditure category. The term segmentation criterion relates to the nature of the information used for market segmentation and can also relate to one specific construct, such as benefits sought.

Common Segmentation Criteria

The most common segmentation criteria are geographic, socio-demographic, psychographic, and behavioral.

1. **Geographic Segmentation**: Typically uses the consumer's location of residence as the criterion to form market segments. This approach is simple and often appropriate for certain products and services. For example, language differences across countries can represent a pragmatic reason for geographic segmentation.
2. **Socio-Demographic Segmentation**: Includes criteria such as age, gender, income, and education. Socio-demographic segments can be useful in industries like luxury goods, cosmetics, baby products, and retirement villages. These criteria are easily determined for each consumer but may not always explain product preferences.
3. **Psychographic Segmentation**: Groups people according to psychological criteria, such as beliefs, interests, preferences, aspirations, or benefits sought when purchasing a product. Psychographic criteria are more complex than geographic or socio-demographic criteria but are more reflective of the underlying reasons for differences in consumer behavior.
4. **Behavioral Segmentation**: Searches for similarities in behavior or reported behavior, such as prior experience with the product, frequency of purchase, amount spent, and information search behavior. Behavioral segmentation is advantageous as it uses actual behavior as the basis of segment extraction, but behavioral data is not always readily available, especially for potential customers who have not previously purchased the product.

Summary

Good market segmentation analysis requires high-quality empirical data. Organizations must carefully choose segmentation criteria that best fit their marketing context, balancing simplicity with the need for meaningful insights into consumer behavior. Geographic and socio-demographic criteria are straightforward but may lack depth, while psychographic and behavioral criteria offer richer insights at the cost of increased complexity.

Step 4: Exploring Data

First Glimpse at the Data

- **Initial Exploration:**
  - Understand data structure using summary statistics and visualizations.
  - Identify any immediate issues or patterns in the data.

Pre-Processing

1. **Categorical Variables:**
   - **Merging Levels:**
     - Merge categories if original ones are too granular (e.g., income ranges).
     - Example: Merging higher income categories to balance frequencies:
   - **Converting to Numeric:**

        - Convert ordinal data to numeric if distances between categories are assumed equal.
        - Binary variables can be converted to 0/1 for further analysis:
2. **<u>Numeric Variables:</u>**
    - **Standardization:**
        - Standardize variables to put them on a common scale, especially when ranges differ.
        - Default method: Subtract mean and divide by standard deviation:
    - **Alternative Methods:**
        - Use robust estimates (median, interquartile range) for data with <u>outliers.</u>

<u>Principal Components Analysis (PCA)</u>

1. **Purpose:**
    - PCA transforms multivariate data into new, uncorrelated variables (principal components).
    - The first principal component captures the most variability, followed by the subsequent components.
2. **Process:**
    - PCA operates on covariance or correlation matrices.
    - Use the correlation matrix if the data ranges of variables differ significantly.
3. **Example with Travel Motives Data:**
    - Perform PCA:
    - Inspect the PCA object for standard deviations and rotation matrix.
4. **Output Interpretation:**
    - **Standard Deviations:** Reflect the importance of each principal component.
    - **Rotation Matrix:** Shows how original variables contribute to principal components.
    - Example output:
5. **<u>Variance Explained</u>:**
    - **Proportion of Variance:** Indicates how much of the original data's variance is captured by each principal component.
    - **Cumulative Proportion:** Shows the cumulative variance explained by the components.
    - Example summary:
6. **<u>Visualization:</u>**
    - Plot data in two dimensions using selected principal components (PC2 and PC3 for better differentiation):
7. **<u>Dimensionality Reduction:</u>**
    - PCA can identify redundant variables by showing high loadings on the same components.
    - Using PCA for exploratory analysis is recommended, but using a subset of principal components for segmentation is not advised due to the risk of missing important information.

<u>Key Takeaways</u>

- Pre-process categorical and numeric variables appropriately before analysis.

- Use PCA to transform and explore multivariate data, identifying key components and relationships.
- Standardize data if variable ranges differ significantly.
- Interpret PCA results to understand variance explained by each component and visualize data effectively.

Step 8: Selecting the Target Segment(s)

The Targeting Decision

- **Importance of Targeting:**
  - Selecting target segments is a crucial, long-term decision affecting an organization's future performance.
  - By Step 8, several segments have been profiled and described; now one or more segments must be selected for targeting.
- **Knock-Out Criteria:**
  - Segments must meet criteria like size, homogeneity, distinctness, identifiability, reachability, and the ability of the organization to meet segment needs.
  - It should have ensured that remaining segments comply with these criteria.

Market Segment Evaluation

- **Decision Matrix:**
  - Use decision matrices to visualize and evaluate relative segment attractiveness and organizational competitiveness.
  - Examples include the Boston Matrix and the General Electric/McKinsey Matrix.
- **Evaluation Dimensions:**
  - **Segment Attractiveness:** How attractive is the segment to the organization?
  - **Organizational Competitiveness:** How attractive is the organization to the segment?
- **Generic Segment Evaluation Plot:**
  - X-axis: Segment attractiveness to the organization.
  - Y-axis: Organizational competitiveness for the segment.
  - Circle size: Another criterion relevant to segment selection, like profit potential or loyalty.

Steps to Create Segment Evaluation Plot

1. **Assign Values:**
   - Based on profiles and descriptions from Steps 6 and 7, assign values (1-10) to each segment for each attractiveness criterion.
   - Multiply the criterion weights (from Step 2) with these values to get a weighted value.
2. **Calculate Overall Attractiveness:**
   - Sum the weighted values to get each segment's overall attractiveness.

    - Repeat the process for organizational competitiveness.
3. **Plot the Values:**
    - Use a matrix to store segment attractiveness and competitiveness data.
    - Create the segment evaluation plot

Decision-Making Insights

- **Interpretation of Segment Evaluation Plot**
    - Segments with high attractiveness and high organizational competitiveness are ideal targets (e.g., Segment 8).
    - Segments with high profit potential but low organizational competitiveness may not be suitable (e.g., Segment 5).

Checklist

1. **Meeting:**
    - Convene segmentation team.
2. **Segment Consideration:**
    - Determine potential target segments from those profiled.
3. **Compliance Check:**
    - Ensure all considered segments meet knock-out criteria.
4. **Assign Values:**
    - Agree on values for each segment for attractiveness and competitiveness criteria.
5. **Calculate Attractiveness and Competitiveness:**
    - Multiply segment values with criterion weights and sum up for each segment.
6. **Plot Values:**
    - Create segment evaluation plot.
7. **Preliminary Selection:**
    - Make a preliminary selection of target segments.
8. **Compatibility Check:**
    - Ensure selected segments are compatible if targeting multiple segments.
9. **Advisory Committee:**
    - Present selected segments for discussion and reconsideration if needed.

Prajakta Kulkarni

Step1

Step2

Step3

Step7

SUMMARY – Market Segmentation

Step 1: Deciding (not) to Segment

1. **Long-term Commitment** : Market segmentation necessitates a persistentcommitment and a readiness to make big organizational adjustments, such as creatingnew products, changing old ones, and changing tactics for pricing, distribution, and communication.
2. **Expenses** : Developing a segmentation strategy calls for spending money on surveys, focus groups, research, and producing a variety of marketing collateral. It is imperative that these costs are justified by the anticipated growth in revenue.
3. **Organizational Structure** : In order to optimize the benefits, companies should organize themselves around market segments as opposed to products. This may mean making modifications to internal structures in order to concentrate on the unique requirements of various market segments.
4. **Top-Level Decision:** Senior executives are responsible for deciding whether to undertake market segmentation and for communicating that decision to all other organizational levels and departments.

Implementation Difficulties:

1. **Leadership Concerns:** Strong leadership, proactive advocating, and dedication from senior management are necessary for successful implementation. The process may be compromised by inadequate resources and a lack of senior management support.
2. **Organizational Culture :** Implementation success may be hampered by office politics, a lack of market orientation, resistance to change, and poor communication.
3. **Lack of Training and Expertise:** Failure might result from senior management and the segmentation team's lack of knowledge about market segmentation. Another obstacle may be the lack of an established marketing department or skilled specialists.
4. **Objective Restrictions:** The organization's capacity to carry out the required adjustments may be hampered by a lack of funding and structural constraints.

Step 1 Checklist:

1. **Evaluate Market Orientation:** Make sure the company is open to new ideas, flexible, and focused on the market.
2. **Assess Communication and Resources:** Adequate financial resources and effective communication between units are essential.
3. **Secure Commitment**: Get senior management's active, visible commitment, along with funding.
4. **Training and Understanding:** Through training, make sure that everyone in the company has a solid understanding of market segmentation and its ramifications.

5. **Team Assembly:** Assemble a focused segmentation team comprising data and marketing professionals.
6. **Organized Process:** Create an organized procedure with well-defined goals, delegated tasks, and enough time to perform the analysis without feeling rushed.

If these criteria are not met, it may be wise to reconsider pursuing a market segmentation strategy.

Step 2: Specifying the Ideal Target Segment:

1. **Segment Assessment Standards**
   Businesses need to actively participate in market segmentation study, particularly in Step 2, which lays the groundwork for subsequent phases. This entails establishing two sets of criteria: attractiveness criteria (used to assess the relative appeal of segments) and knock-out criteria (important, non-negotiable attributes for targeting). Numerous segment evaluation criteria are available in the literature and can be customized to meet the demands of the company.

2. ### Distinctive Features
   The use of knock-out criteria aids in identifying the market categories that warrant additional analysis. Which are homogeneity , size , Identifiability , reachability and organizational match.

3. ### Criteria for Attractiveness
   Each market category is rated according to attractiveness criteria. In contrast to knockout criteria, which evaluate segments on a continuum, beauty criteria are not binary. In Step 8, the segments that are chosen as target segments are determined by their overall attractiveness.

4. ### Putting a Structured Process in Place
   Segments can be assessed according to their organizational competitiveness and aesthetic appeal with the aid of an organized procedure, including the use of a segment evaluation plot. A varied team with members from different organizational units should negotiate and decide on the criteria's factors. This early choice of attractiveness standards facilitates the target segment selection process in Step 8 and guarantees thorough data gathering in Step 3.

Step 2 Checklist

1. Convene a segmentation team meeting.
2. Agree on knock-out criteria (homogeneity, distinctness, size, match, identifiability, reachability).
3. Present knock-out criteria to the advisory committee for review.
4. Review available criteria for segment attractiveness.
5. Agree on a subset of up to six attractiveness criteria with the segmentation team.

6. Distribute 100 points among the selected attractiveness criteria to reflect their relative importance.
7. Discuss and agree on the weightings with the segmentation team.
8. Present the criteria and weightings to the advisory committee for review and adjustment if necessary.

Step 3: Collecting Data:

Segmentation Variables

Market segmentation is covered throughout the literature via both data-driven and commonsense methods. Commonsense segmentation divides consumers into segments based on a single variable (gender, for example), and then uses descriptor variables—other personal characteristics—to describe each segment. On the other hand, data-driven segmentation makes use of a variety of characteristics (such vacation benefits) to produce segments that accurately represent either artificially designed or naturally occurring groupings. In order to precisely classify customers into groups and to properly describe them, which informs marketing tactics like product customisation, pricing, distribution, and communication, the quality of empirical data is essential.

Segmentation Criteria

The crucial choice that businesses must make when deciding on a segmentation criterion prior to data collection is highlighted in the text. These criteria, which include behavioral, psychographic, demographic, and geographic aspects, influence the way in which consumers are categorized for marketing objectives. For instance, geographic segmentation splits markets according to geography, making focused communication easier to implement but sometimes omitting other aspects of the consumer. Despite difficulties like cultural biases in survey data, geographic segmentation is useful despite its simplicity, especially in global environments.

Socio-Demographic Segmentation

Socio-demographic segmentation focuses on age, gender, income, and education, widely used in industries like luxury goods, cosmetics, baby products, retirement villages, and tourism. While these criteria simplify segment identification and sometimes explain specific product preferences (e.g., family vacation choices based on having children), they typically explain only about 5% of consumer behaviour variance. Values, tastes, and preferences are considered more influential for segmentation decisions than socio-demographic factors.

Psychographic segmentation

Psychographic segmentation groups people by psychological criteria like beliefs, interests, preferences, aspirations, or benefits sought. It includes benefit segmentation (Haley, 1968) and lifestyle segmentation (Cahill, 2006). This approach is more complex than geographic or socio-demographic segmentation and often uses multiple variables. It effectively reflects underlying reasons for consumer behavior but is complex to determine segment memberships and relies on reliable and valid measures.

Behavioural segmentation

Behavioural segmentation groups people by similarities in behavior like purchase frequency or amount spent. Advantages include using actual behavior for direct segment extraction, such as expenses (Tsai and Chiu, 2004) and brand choice (Poulsen, 1990). Challenges include unavailable behavioral data for new customers.

Survey Data :

Common Use: Widely used in market segmentation.

Advantages: Easy and cheap to collect.

Biases: Can be biased, affecting-quality.

Considerations :

Choice of Variables: Only include relevant ones to reduce noise and respondent fatigue.

Response Options: Binary or metric are preferable; ordinal scales can complicate analysis.

Response Styles: Biases like extreme answers can mislead results.

Sample Size: Larger samples improve segment accuracy; at least 60 times the segmentation variables (Dolnicar et al., 2014).

Internal Sources:

Examples: Scanner data, booking data.

Advantages: Reflect actual behavior, unbiased by memory or responses.

Challenges: May over-represent existing customers, missing new customer data.

Experimental Studies:

Examples: Responses to ads, choice experiments.

Usage: Understand how product_attributes affect choices.

Step 7: Describing Segments :

Developing a Complete Picture of Market Segments

Developing a complete picture of market segments involves segment profiling to understand differences in segmentation variables chosen early in the process. This profiling is similar to the step of describing segments but uses additional information not used in extraction. Profiling segments is like dating to know a potential spouse before marriage, ensuring no surprises later. For example, in a data-driven market segmentation analysis using Australian travel motives data, profiling investigates segment differences in travel motives. Descriptions use additional variables like age, gender, past travel behavior, and media use. Detailed segment descriptions are crucial for developing a customized marketing mix. Visualizations and statistical analyses help study these differences effectively.

Using Visualisations to Describe Market Segments

Using graphical statistics to describe market segments offers two main advantages: it simplifies result interpretation for both analysts and users, and integrates information on the statistical significance of differences, thus avoiding the over-interpretation of insignificant variations. Graphical representations effectively convey the essence of marketing research results and are preferred by managers for their intuitiveness. Various charts are suitable for visualizing differences in nominal, ordinal, and metric descriptor variables, such as gender, education level, age, and spending.

Nominal and Ordinal Descriptor Variables

When describing differences between market segments using nominal or ordinal descriptor variables, a cross-tabulation of segment membership with the descriptor variable is essential. For example, in the Australian travel motives data set, segment membership can be visualized and statistically tested. Adding segment membership as a categorical variable to the descriptor variables data frame allows for generating cross-tabulations. For instance, a cross-tabulation of gender and segment membership can be visualized using stacked bar charts or mosaic plots. While stacked bar charts show segment sizes, mosaic plots provide a clearer comparison of proportions and incorporate inferential statistics, using colours to indicate significant differences. Mosaic plots can also handle multiple descriptor variables, highlighting associations and aiding in interpreting segment characteristics.

Metric Descriptor Variables

R packages like lattice and ggplot2 offer conditional plots that visualize differences between market segments using metric descriptor variables. Conditional plots, divided into panels or facets, display subsets of data, such as different market segments. For example, segment age distribution and moral obligation scores can be visualized using histograms. However, additional insights are gained through parallel box-and-whisker plots, which show distribution differences more clearly. These plots can incorporate statistical elements like confidence intervals, enhancing interpretation. Mosaic plots and modified segment-level stability across solutions (SLSA) plots also visualize segment differences and stability, using colors to represent descriptor values and highlight significant associations.

Testing for Segment Differences in Descriptor Variables

Simple statistical tests are crucial for formally assessing differences in descriptor variables across market segments. Segment membership, akin to a nominal variable, can be evaluated for associations with other nominal or ordinal variables using methods such as the chi-squared (χ^2) test, which examines independence in contingency tables. For instance, to test if there are significant differences in gender distribution across segments, a chi-squared test can be employed. Similarly, metric variables such as age or spending can be analyzed using ANOVA, which compares means across multiple segments. These tests provide statistical evidence of whether segment differences are significant and help in understanding the unique characteristics of each market segment based on the variables of interest.

Binary Logistic Regression

Model Setup: Used glm() with binomial family and logit link to predict membership in segment 3 based on age and moral obligation scores.

Interpretation: Coefficients (Intercept, Age, Obligation2Q2-Q4) show how log odds of segment 3 membership change with age and moral obligation categories.

Evaluation: Assessed model fit with deviance, AIC, and significance tests (z value, Pr(>|z|)).

Visualization: Used effects package for plots showing how predicted probabilities vary with age and moral obligation.

Model Selection: Employed step() function to select variables (Education, NEP, Vacation.Behaviour) based on AIC for improved predictive accuracy.

Prediction: Compared predicted probabilities of segment membership using predict() and visualized with boxplots to assess model performance.

Multinomial Logistic Regression

Multinomial logistic regression in R, specifically using the multinom() function from the nnet package, is suitable when predicting categorical segments (more than two). It assumes a multinomial distribution with a logistic link function.

The model estimates coefficients for each segment (except the baseline) based on predictors like Age and Oblig2, indicating how log odds change across segments. Model evaluation involves interpreting these coefficients and their significance in predicting segment membership.

Key steps include assessing model fit using deviance and AIC, understanding variable importance through tests like Anova(), and visualizing predicted probabilities to compare with observed data. Effect plots illustrate how predictors influence segment probabilities, aiding interpretation of model predictions.

Tree-Based Methods

Tree-based methods, such as classification and regression trees (CART), offer a supervised learning approach for predicting categorical or binary outcomes using independent variables. These methods excel in variable selection, are easy to interpret through visual representations of decision trees, and can incorporate interaction effects straightforwardly. They are robust with large numbers of predictors but may produce unstable results due to sensitivity to minor changes in data.

The approach involves recursively partitioning the dataset into groups that are as homogeneous as possible with respect to the outcome variable. Each split in the tree aims to maximize the purity of resulting groups. Terminal nodes represent the final segments where predictions are made based on majority outcomes within each node. Different algorithms vary in their splitting criteria, stopping rules, and methods for final predictions.

Packages like rpart and partykit in R implement these algorithms. rpart follows the original CART method, while partykit offers alternatives with unbiased variable selection. Visualizing trees with tools like plot() provides insights into how variables influence predictions, showing splits and terminal nodes' predictive accuracy for each segment.

Step 7 Checklist :

Bring across Segmentation Solutions: Transfer one or a few chosen segmentation solutions from Step 6 based on appealing profiles.

Select Descriptor Variables: Choose additional consumer information not used in segment extraction, relevant for segmentation analysis.

Visualize Segment Differences: Utilize appropriate visualizations like mosaic plots for categorical/ordinal variables and box-and-whisker plots for metric variables to understand segment variations.

Test Descriptor Variables: Assess statistical significance of descriptor variables. Correct for multiple testing if multiple tests are used.

Introduce Segments to Team: Present each market segment to team members to evaluate understanding. Seek additional insights where needed for comprehensive segment understanding.

Github links

Pratiksha Gaikwad:

https://github.com/Pratiksha858/Market-segment-analysis

B Pavan Srinivasarao:

https://colab.research.google.com/drive/1GotVfLKhuUej4HUDtd49TxcgTvJZsYQ3?usp=sharing

Himesh Ray:

https://github.com/himesh992/Himesh992

Prajakta Kulkarni:

https://github.com/prajakta0515/Market-segmentation-visualisation