Project Study Task by Kavita Lamani

Market Segmentation

Step 1: Deciding (not) to Segment:

Summary:

Before committing to market segmentation, organizations must understand the implications and barriers associated with this strategy. Market segmentation requires long-term commitment and significant changes, including product development, pricing adjustments, and organizational restructuring. Senior management buy-in and involvement are crucial, along with a market-oriented culture, openness to change, and sufficient resources.

Implications of Committing to Market Segmentation:

1. Market segmentation requires long-term commitment and substantial investments.

2. Organizational changes may be necessary, including product modifications, pricing adjustments, and changes in communication strategies.

3. Strategic business units focused on market segments facilitate on-going market alignment.

Implementation Barriers:

1. Lack of senior management leadership, commitment, and resource allocation.

2. Organizational culture barriers such as resistance to change, short-term thinking, and lack of market orientation.

3. Lack of training and understanding of market segmentation concepts.

4. Absence of formal marketing function or qualified marketing expertise.

5. Objective restrictions such as financial constraints or inability to make necessary structural changes.

6. Process-related barriers like unclear objectives, lack of planning, and time pressure.

Step 1 Checklist:

1. Assess organizational culture, willingness to change, and long-term perspective.

2. Secure visible commitment and involvement from senior management.

3. Ensure understanding of market segmentation concepts and implications.

4. Establish a dedicated segmentation team with necessary expertise.

5. Develop a structured process for market segmentation analysis.

6. Clarify objectives, assign responsibilities, and ensure adequate time for analysis.

.

Step 2: Specifying the Ideal Target Segment:

- Step 2 of market segmentation involves crucial input from the organization, with user involvement spanning various stages of analysis.
- Two sets of segment evaluation criteria are essential: knock-out criteria and attractiveness criteria.
- Knock-out criteria determine if segments qualify for further assessment, focusing on aspects like homogeneity, distinctness, size, match, identifiability, and reachability.
- Attractiveness criteria, more diverse and detailed, help evaluate the relative attractiveness of remaining segments, covering factors like size, growth, profitability, accessibility, and compatibility with the company.
- A structured process is recommended for segment evaluation, often utilizing segment evaluation plots.
- The segmentation team, ideally comprising representatives from different organizational units, should select around six segment attractiveness criteria and assign weights to each based on relative importance.
- Approval from the advisory committee, representing various perspectives, is sought for finalizing segment attractiveness criteria and their weights.

Step 3: Collecting Data:

- Segmentation variables are the foundation of both common sense and data-driven market segmentation, derived from empirical data to define market segments.
- In common sense segmentation, a single characteristic (e.g., gender) is used to split the sample into segments, while other personal characteristics serve as descriptor variables to detail these segments.
- Data-driven segmentation utilizes multiple segmentation variables to identify or create market segments, providing a more nuanced understanding of consumer behaviour and preferences.
- Market segmentation relies on empirical data to identify or create distinct market segments based on consumer characteristics and behaviours.
- In data-driven segmentation, multiple variables are utilized to define segments based on commonalities in consumer preferences or behaviours.
- Segmentation variables can range from demographic factors like gender and age to psychographic factors such as preferences and aspirations.
- Descriptor variables provide additional detail about segments and can include socio-demographic information or media behaviour.
- The choice of segmentation criterion depends on the organization's knowledge of the market and the suitability of criteria like geographic, socio-demographic, psychographic, or behavioural factors.
- Geographic segmentation considers consumers' location, facilitating targeted communication but may overlook other relevant characteristics.
- Socio-demographic segmentation based on age, gender, income, and education offers insights into consumer preferences in specific industries but may not always explain product preferences adequately.
- Psychographic segmentation delves into consumers' beliefs, interests, and benefits sought, providing a deeper understanding of consumer behaviour but requiring more complex analysis.

- Behavioural segmentation focuses on actual behaviours or reported behaviour patterns, offering insights into consumer actions but may require extensive data collection.
- Survey data is a common source for segmentation studies but can be influenced by biases, such as response styles and sampling errors.
- Careful selection of variables and response options is crucial to ensure the quality of data for segmentation analysis.
- Sample size plays a vital role in segmentation analysis, with recommendations suggesting a sufficient number of respondents relative to the number of segmentation variables.
- High-quality data free of biases and with adequate sample size is essential for accurate segmentation analysis and effective marketing strategies.
- Internal data from organizations, such as scanner data from grocery stores or booking data from airline loyalty programs provide valuable insights into actual consumer behaviour.
- The strength of internal data lies in its representation of real consumer actions rather than self-reported behaviour, which can be influenced by memory biases and response biases.
- Internal data is often automatically generated and easily accessible if stored properly, requiring minimal effort for collection.
- However, using internal data may lead to biased insights by over-representing existing customers, thus potentially missing out on information about potential future customers with different consumption patterns.
- Experimental data, obtained from field or laboratory experiments, can also serve as a source for segmentation analysis.
- Experimental studies, such as that involving consumer response to advertisements or choice experiments, provide valuable insights into consumer preferences and behaviours that can be used as segmentation criteria.

Step 7: Describing Segments:

Developing a Complete Picture of Market Segments:

- Market segment profiling involves understanding differences in segmentation variables across segments, which are chosen early in the segmentation analysis process.
- Step 7, describing segments, is similar to profiling but involves using additional information about segment members, such as demographic, psychographic, and behavioral variables.
- Profiling investigates differences between segments regarding segmentation variables, while segment description uses descriptor variables like age, gender, past behavior, media use, and spending patterns.
- Detailed segment descriptions are crucial for developing a customized marketing mix tailored to each segment's characteristics.
- For example, knowing that segment 4 of the Australian travel motives dataset cares about nature is insightful, but understanding their demographic details, behavior patterns, and media preferences enables targeted marketing strategies.
- Differences between segments regarding descriptor variables can be studied through descriptive statistics, visualizations, or inferential statistics.
- While traditional marketing literature often uses statistical testing and tabular presentations, visualizations can enhance the user-friendliness of segment descriptions.

Using Visualisations to Describe Market Segments:

- Visualizations are crucial for describing differences in market segments based on descriptor variables, which can be either nominal/ordinal (e.g., gender, education level) or metric (e.g., age, spending).
- They simplify interpretation for both analysts and users and integrate information on statistical significance, preventing over-interpretation of insignificant differences.

- For nominal/ordinal descriptor variables, cross-tabulations are often used as the basis for visualizations. Stacked bar charts and mosaic plots are common choices, with mosaic plots being particularly useful for comparing proportions across segments.
- Mosaic plots can incorporate inferential statistics, highlighting differences between observed and expected frequencies using colour coding.
- For metric descriptor variables, histograms and box-and-whisker plots are commonly used. Conditional plots, such as histograms by segment or boxplots for each segment, provide comparative views.
- Box-and-whisker plots can include elements of statistical inference, such as confidence intervals for medians, aiding in the interpretation of differences between segments.
- Segment stability across different segmentation solutions can be visualized using modified segment level stability across solutions (SLSA) plots, where node colors represent mean values of metric descriptor variables, providing insights into consistent segment characteristics across solutions.

Testing for Segment Differences in Descriptor Variables:

The process of testing for segment differences in descriptor variables involves several steps and statistical methods. Here's a summary of the key points:

1. Testing for Nominal/Ordinal Variables:

- Segment membership, representing a nominal summary statistic of segmentation variables, can be tested for association with other nominal or ordinal variables.
- Cross-tabulation using mosaic plots help visualize associations, and the $\chi 2$-test is used to test for independence.
- For example, the $\chi 2$-test can be used to test for significant differences in gender distribution across market segments.

2. Testing for Metric Variables:

- Associations between segment membership and metric variables are visualized using parallel boxplots.
- Analysis of Variance (ANOVA) is commonly used to test for significant differences in means across multiple segments.
- Pairwise comparisons between segments provide additional insights into which segments differ significantly.
- Tukey's honest significant differences can be plotted to visualize pairwise segment comparisons.

3. Adjusting for Multiple Testing:

- P-values need to be adjusted for multiple testing to control the overall Type I error rate.
- Holm's method is suggested as a less conservative approach compared to Bonferroni correction.
- Other methods such as the false discovery rate procedure are also available.

4. Alternative Approaches:

- Instead of pairwise t-tests, Tukey's honest significant differences can be plotted to visualize segment differences.

- Confidence intervals of differences in mean values are displayed, with significant differences indicated by intervals not crossing zero.

5. Interpreting Results:

- Visualizations like parallel boxplots and Tukey's honest significant differences provide insights into the direction and significance of segment differences.
- Segments can be characterized based on their distinct behaviors or characteristics revealed through statistical testing.

Predicting Segments from Descriptor Variables:

- **Regression Model for Market Segments Prediction:**
  - Regression models are utilized to predict segment membership using descriptor variables.
  - The dependent variable is the segment membership, which is categorical, and the independent variables are descriptor variables.
  - The regression model aims to find a function that predicts segment membership based on the descriptor variables.
- **Linear Regression Model:**
  - The basic regression model used is the linear regression model.
  - It assumes a linear relationship between the dependent variable and the independent variables.
  - The model estimates coefficients ($\beta 0$, $\beta 1$, ..., βp) that represent the mean age difference between segments.
- **Regression Coefficients Interpretation:**
  - Each coefficient represents the mean difference in the dependent variable (age) between segments when other variables are held constant.
  - The intercept ($\beta 0$) represents the mean age of one of the segments, typically chosen as a reference category.
- **Generalized Linear Models (GLMs):**
  - GLMs are introduced as a more flexible framework for regression analysis.
  - They can accommodate a wider range of distributions for the dependent variable, such as categorical variables.
  - GLMs use a link function to transform the mean value of the dependent variable to an unlimited range.
- **Special Cases of GLMs:**
  - Binary and multinomial logistic regression are discussed as special cases of GLMs.
  - These models are used when the dependent variable follows either a binary or a multinomial distribution, respectively.
  - The link function used in these models is the logit function.
  - Overall, regression analysis, particularly linear regression and its extensions like GLMs, provides a framework for predicting market segment membership based on descriptor variables and understanding the critical variables that influence segment identification.

Binary Logistic Regression:

1. Model Fitting:

- Binary logistic regression models are fitted using the `glm()` function in R, specifying the family argument as `binomial()` for the Bernoulli distribution with a logit link function.
- The formula interface (`~`) is used to specify the relationship between the dependent variable and independent variables.

2. Interpretation of Coefficients:

- The coefficients represent the log odds of the dependent variable (in this case, belonging to segment 3) with respect to the independent variables.
- The intercept gives the log odds when all independent variables are zero.
- Interpretation of coefficients involves transforming them with the inverse link function (inverse logit function) to obtain predicted probabilities.

3. Visualizing Effects:

- The effects of independent variables on the predicted probabilities can be visualized using the `effects` package in R.
- These plots show how predicted probabilities change with different values of independent variables.

4. Model Evaluation:

- Model fit is evaluated using deviance, AIC, and significance tests for coefficients.
- Significance tests assess whether coefficients significantly affect the dependent variable.
- The Anova function from the `car` package can be used to compare nested models and assess variable significance.

5. Model Selection:

- Overfitting is addressed by selecting relevant independent variables using methods like stepwise selection based on AIC.
- Stepwise selection iteratively adds or removes variables to improve model fit until no further improvement is achieved.

6. Predictive Performance:

- Predictive performance of the model is evaluated using predicted probabilities for segment membership.
- Boxplots can visually compare the predicted probabilities between different models to assess their discriminative ability.

Predicting Segments from Descriptor Variables:

1. Binary Logistic Regression:

- Formulates a regression model for binary data using the Bernoulli distribution and the logit link function.
- Uses the `glm()` function in R to fit logistic regression models, specifying the family as binomial with the logit link.
- Interprets coefficients as changes in log odds, indicating the effect of independent variables on the dependent variable.
- Predicts probabilities of binary outcomes using the inverse logit function.
- Evaluates model fit using deviance, AIC, and significance tests for coefficients.
- Utilizes package effects in R for visualizing predicted probabilities for different levels of independent variables.
- Performs model selection using stepwise procedures based on AIC to improve predictive performance.

2. Classification Trees:

- Constructed using recursive partitioning to split data into homogeneous groups based on independent variables.
- Implemented in R using the `ctree()` function from the `partykit` package.
- Splitting nodes based on variables that maximize homogeneity within resulting groups.
- Terminal nodes represent final predictions, with proportions indicating class membership.
- Visualized using tree plots, illustrating splits and terminal nodes with associated class proportions.
- Can handle categorical dependent variables with multiple levels, facilitating segmentation analysis.

3. Comparison:

- Logistic regression provides interpretable coefficients but assumes linear relationships.
- Classification trees offer flexibility in handling non-linear relationships and interactions but may be less interpretable.
- Model selection in logistic regression relies on AIC, while tree algorithms automatically handle variable selection.
- Performance evaluation involves assessing predictive accuracy and interpretability.

4. Considerations:

- Logistic regression is suitable for understanding the impact of continuous and categorical predictors on binary outcomes.
- Classification trees are useful for exploring complex interactions and non-linear relationships in the data.

- Both methods require careful consideration of model assumptions, such as linearity and independence of observations.
- Model selection should balance predictive performance and model complexity to avoid overfitting.