MARKET SEGMENT IMPLICATION SUMMARY(TASK 2)

By Uma Maheshwari M(Team Lead), Mriganka, Divyanshu and Meenakshi

Introduction

Making a plan for a company's marketing initiatives is called market planning. It entails examining market trends, determining target audiences, and creating engagement and outreach tactics for those consumers.

A marketing strategy is composed of two parts:

1. Strategic
2. Tactical

The strategic plan explains the organization's long-term course, but it does not go into great depth about the immediate marketing steps that must be taken to carry out this course. Contrarily, the tactical marketing plan operates. It converts the lengthy strategic strategy into comprehensive guidelines for immediate marketing action. To sum up, strategic and tactical marketing are equally important for an organization to succeed. Poor strategic marketing will never be made up for by effective tactical marketing. The cornerstone of an effective organization is strategic marketing.

Market Segment Analysis

The technique of splitting a bigger market into smaller, more focused groups of consumers who have comparable traits and demands is known as market segmentation. As a result, businesses are better able to identify and target their ideal clients and adjust their marketing strategies to better suit their individual requirements. Market segmentation is a crucial component of strategic marketing.

Step-by-Step Analysis of Market Segment

Step1: Deciding (not) to Segment

Market segmentation, while being a key marketing strategy used in many organizations, may not always be the best decision due to the implications and commitment required. The organization needs to commit to the segmentation strategy in the long term, as it requires substantial changes and investments. This includes potentially developing new products, modifying existing products, changing pricing and distribution channels, and adjusting the internal structure of the organization. The decision to pursue market segmentation should be made at the highest executive level and systematically communicated and reinforced across all organizational levels and units.

There are several barriers to successfully implementing a market segmentation strategy. These include lack of leadership, pro-active championing, commitment, and involvement from senior management, as well as inadequate resources allocated for market segmentation analysis and implementation. Organizational culture can also be a barrier, including lack of market or consumer orientation, resistance to change, lack of creative thinking, poor communication and sharing of information, short-term thinking, unwillingness to make changes, and office politics. Lack of training, formal marketing function or qualified marketing experts, and data management and analysis can also hinder successful implementation. Objective restrictions, such as limited financial resources or structural changes required, can also pose obstacles. Process-related barriers may include unclear objectives, lack of planning or structured processes, lack of allocation of responsibilities, and time pressure.

To overcome these barriers, market segmentation analysis should be made easy to understand and results should be presented in a clear and concise manner. Proper planning, allocation of resources, and involvement of senior management are crucial for successful implementation. Organizational culture should be aligned with market orientation and a qualified marketing expert should be involved. Process-related challenges can be addressed by establishing structured processes and allocating responsibilities. Overall, organizations need to carefully consider the implications and challenges of market segmentation before deciding to pursue this strategy.

The checklist provided includes tasks and questions to determine if an organization is ready to implement a market segmentation strategy.
The checklist includes the following:
Determine if the organization's culture is market-oriented. If the answer is yes, proceed. If the answer is no, seriously consider not proceeding.
Determine if the organization is genuinely willing to change. If the answer is yes, proceed. If the answer is no, seriously consider not proceeding. Determine if the organization takes a long-term perspective. If the answer is yes, proceed. If the answer is no, seriously consider not proceeding.

Step2: Specifying the Ideal Target Segment

The third layer of market segmentation analysis depends primarily on user input, and it is important for the user to be involved in most stages of the analysis process. This involvement should go beyond just providing a briefing at the start or developing a marketing mix at the end. After committing to investigating the value of a segmentation strategy in Step 1, the organization needs to make a major contribution to market segmentation analysis in Step 2, which involves determining two sets of segment evaluation criteria.

The first set of evaluation criteria, referred to as knock-out criteria, are the essential and non-negotiable features of segments that the organization would consider targeting. The second set of evaluation criteria, referred to as attractiveness criteria, are used to evaluate the relative attractiveness of the remaining market segments that comply with the knock-out criteria.

Various criteria for evaluating market segments have been proposed in the literature, including factors such as size, growth, accessibility, profitability, compatibility with the company, competitive advantage, technological factors, socio-political factors, and more. These criteria are described at different levels of detail and are proposed by different authors in chronological order. Some authors also emphasize the importance of factors such as consumer motivation, competitor's ability and motivation to retaliate, company's current position in the market, and potential for creating value for customers.

Overall, the involvement of the user in the market segmentation analysis process, including the determination of evaluation criteria, is crucial for producing results that are useful to the organization.

Step 3: Collecting Data

Segmentation variables are used in market segmentation to divide a sample of consumers into distinct segments. Empirical data, which can come from sources such as surveys, observations, or experimental studies, forms the basis of segmentation. In commonsense segmentation, a single characteristic of consumers, such as gender, is used as the segmentation variable to create segments. Other personal characteristics, known as descriptor variables, are used to describe the segments in detail.

Data-driven market segmentation, on the other hand, uses multiple segmentation variables to identify or create segments based on shared characteristics or behaviors. The quality of empirical data is crucial for developing valid segmentation solutions, as it affects the accuracy of segment assignments and descriptions.

Common segmentation criteria include geographic, socio-demographic, psychographic, and behavioral variables. The choice of segmentation criterion should be based on prior knowledge about the market and the marketing context. It is generally recommended to use the simplest possible approach for segmentation.

The presence of noisy variables in market segmentation analysis can hinder the ability of algorithms to identify correct market segments. Noisy variables can result from poorly developed survey questions or redundant variables in the survey items. To avoid this issue, it is recommended to carefully develop survey questions, include only necessary and unique questions, and resist the temptation to include redundant questions.

The response options provided to respondents in surveys also play a crucial role in the quality of data for segmentation analysis. Binary or metric response options are preferred as they allow for distance measures and statistical analysis, while ordinal data can be more complicated to analyze. Visual analogue scales or slider scales can be used to capture fine nuances of responses in a metric format. Binary response options have been shown to outperform ordinal options in many contexts.

Response styles, such as biases in answering survey questions, can also impact segmentation results. Common response styles include using extreme or midpoint options consistently, or agreeing with all statements. These response styles can affect the accuracy of segmentation results as algorithms may not be able to differentiate between true beliefs and response biases.

In conclusion, careful consideration should be given to the development of survey questions, selection of variables, and response options in market segmentation analysis to avoid noisy variables and response biases that can hinder the accuracy of segmentation results. Using binary or metric response options, and being mindful of response styles, can help improve the quality of data for segmentation analysis.

Step 4: Exploratory Analysis(by Uma Maheshwari M)

The process of data exploration involves cleaning and preprocessing data after data collection, and it helps in identifying measurement levels of variables, investigating univariate distributions of variables, and assessing dependency structures between variables. Data may need to be pre-processed and prepared for input into different segmentation algorithms. The results of data exploration provide insights into the suitability of different segmentation methods for extracting market segments.

In this example, the travel motives data set is used to illustrate data exploration. The data set contains 20 travel motives reported by 1000 Australian residents in relation to their last vacation. The data is stored in a data frame named "vac". The data is inspected using commands such as "colnames(vac)" to get the column names and "dim(vac)" to get the dimensions of the data set. A summary of the data set is generated using the "summary(vac)" command, which provides information about the variables Gender, Age, Income, and Income2. The summary includes descriptive statistics such as minimum, maximum, median, mean, and quartiles for metric variables like Age and Income. It also indicates that there are missing values (NAs) in the variables Income and Income2, with 66 respondents not providing income information in the survey.

Data cleaning is an important step before starting data analysis. It involves checking for errors in data collection or data entry, ensuring consistent labels for categorical variables, and checking for implausible values in metric variables. For example, age values should fall within a plausible range, such as 0 to 110 years. Categorical variables should only contain permissible values, and any other values need to be corrected during the data cleaning process.

In the case of the Australian travel motives data set, data cleaning is needed for the variable Income2, as the categories are not sorted in order due to how data is read into R. Factors are the default format for storing categorical variables in R, and their levels are sorted alphabetically. To re-order the categories, the column can be copied to a helper variable, the levels can be stored in a separate variable, the correct re-ordering of levels can be found, and then the variable can be converted into an ordered factor in R.

It is important to check the accuracy of the data transformation before overwriting the original column of the data set. Cross-tabulations can be used to compare the original column with the re-ordered version to ensure that no errors were introduced during the data cleaning process.

Using code for data cleaning, exploration, and analysis ensures reproducibility, as every step can be documented and replicated in future analyses. Descriptive analysis, such as numeric summaries and graphical representations, can provide insights into the data and help with data interpretation. Histograms, box plots, scatter plots, and bar plots are useful graphical methods for visualizing numeric and categorical variables. Histograms can reveal the distribution of numeric variables, while bar plots can show frequency counts for categorical variables. Mosaic plots can be used to illustrate the association of multiple categorical variables.

Pre-processing is an important step in data analysis that involves preparing and cleaning the data before conducting further analysis. This section focuses on pre-processing procedures for categorical and numeric variables.

For categorical variables, there are two common procedures used: merging levels of categorical variables and converting categorical variables to numeric ones. Merging levels of categorical variables is useful when the original categories are too differentiated and result in imbalanced frequencies. For example, if the income categories in a survey dataset have very few respondents in some categories, merging them with adjacent categories with similar frequencies can result in a new variable with more balanced frequencies.

Converting categorical variables to numeric ones can be done when it makes sense to do so. For example, ordinal data, such as income categories, can be converted to numeric data if it can be assumed that the distances between adjacent scale points are approximately equal. Another example is the Likert scale used in consumer surveys, where the assumption is often made that the distances between the answer options are the same. However, it is important to consider the consequences of such conversions, as there may be potential issues with response styles at both the individual and cross-cultural level.

On the other hand, binary variables, which have only two categories, are less prone to capturing response styles and do not require pre-processing. They can easily be converted to numeric variables, and most statistical procedures work correctly after conversion. Dichotomous ordinal or nominal variables can also be converted to binary 0/1 variables without any issues.

Numeric variables, on the other hand, may need to be standardized to put them on a common scale, especially when using distance-based methods for segmentation analysis. Standardizing variables involves transforming them in a way that puts them on a common scale, which helps to balance their influence on segmentation.

Overall, pre-processing is an essential step in data analysis to ensure that the data is clean, prepared, and appropriate for further analysis. This involves handling categorical variables through merging levels or converting them to numeric ones, and standardizing numeric variables to balance their influence on analysis results.

The task involves exploring the data to identify any inconsistencies or systematic contaminations. If needed, data cleaning and preprocessing should be performed. Additionally, the number of segmentation variables should be checked to ensure that there are at least 100 consumers for each variable. If there are too many variables, a subset can be selected using appropriate approaches. Further, the correlation among segmentation variables should be assessed, and a subset of uncorrelated variables should be chosen. Finally, the cleaned and pre-processed data can be passed onto the next step for segment extraction.
Code: :
https://github.com/umamaheshwari20/McDonalds-Case-Study/blob/main/McDonalds%20Code(MSA)%20by%20Uma%20Maheshwari%20M.ipynb

Step 5: Extracting Segments(By Uma Maheshwari & Mriganka)

The process of data-driven market segmentation analysis is exploratory in nature, as consumer data sets are often unstructured and do not contain clear groups of consumers. Consumer preferences are typically spread across the entire plot of a two-dimensional representation of their product preferences. The choice of segmentation method and algorithm used to extract market segments from the data strongly depends on the assumptions made on the structure of the segments implied by the method.

Many segmentation methods used in market segmentation analysis are taken from the field of cluster analysis, where market segments correspond to clusters. Different clustering methods impose different structures on the extracted segments, and it is important to match the data analytic features of the resulting clustering with the context-dependent requirements desired by the researcher. For example, k-means cluster analysis, which aims at finding compact clusters covering a similar range in all dimensions, may fail to identify naturally existing segments with non-compact shapes, such as spirals in the data. On the other hand, single linkage hierarchical clustering, which constructs snake-shaped clusters, may correctly identify such non-compact segments even with an incorrect number of segments specified upfront.

It is important to note that there is no single best algorithm for all data sets, and the choice of algorithm should be guided by the characteristics of the data set and the desired segment characteristics. If the consumer data is well-structured and well-separated, different algorithm tendencies may matter less. However, if the data is not well-structured, the tendency of the algorithm can substantially influence the segmentation solution. Therefore, the algorithm chosen will impose a structure that aligns with the objective function of the algorithm. Understanding the interaction between data and algorithms is critical in the process of grouping consumers into market segments.

This excerpt discusses distance-based methods for finding groups of tourists with similar activity patterns when on vacation, which is a problem of market segmentation. Market segmentation aims at grouping consumers into segments with similar needs or behavior, and in this case, the goal is to group tourists based on their vacation activity preferences. The excerpt introduces the concept of a data matrix, where rows represent observations (tourists) and columns represent variables (vacation activities). Various distance measures are used to quantify the similarity or dissimilarity between two observations.

The three common distance measures discussed in this excerpt are:

Euclidean distance: It is the straight-line distance between two points in a two-dimensional space. It is calculated as the square root of the sum of squared differences between corresponding variables of two vectors.

Manhattan or absolute distance: It gives the distance between two points assuming that streets on a grid (like in Manhattan) need to be used to get from one point to another. It is calculated as the sum of absolute differences between corresponding variables of two vectors.

Asymmetric binary distance: This measure is only applicable to binary vectors where variables can only take values of 0 or 1. It calculates the dissimilarity between two vectors based on the number of dimensions where they share 1s divided by the number of dimensions where at least one of them has a value of 1.

The excerpt also highlights the properties that a distance measure should have, such as symmetry (i.e., the distance between x and y should be the same as the distance between y and x), zero distance for a vector to itself, and compliance with the triangle inequality. It is mentioned that Euclidean distance is the most commonly used distance measure in market segmentation analysis.

The order of the leaves in a tree (such as a dendrogram) is not unique, as at every split into two branches, the left and right branches could be exchanged, resulting in 2n possible dendrograms for the same clustering, where n is the number of observations or consumers in the dataset. This means that dendrograms obtained from different software packages may look different even if they represent the same market segmentation solution. Another source of variation between software packages is how ties are broken, i.e., which two groups are joined first when several have exactly the same distance.

As an example, a dataset on "tourist disasters" is used, which contains survey data collected from adult Australian residents who have undertaken at least one personal holiday in the past 12 months.

The data set has 563 respondents who provided ratings on how often they take risks in six categories. The data is analyzed using hierarchical clustering with Manhattan distance and complete linkage in R. The resulting dendrogram visualizes the sequence of nested partitions and can be cut at a specific height to obtain different market segments.

Hierarchical clustering methods are suitable for small datasets with up to a few hundred observations, but may not be ideal for larger datasets due to the difficulty of interpreting dendrograms and the memory limitations of pairwise distance matrices. In such cases, partitioning clustering algorithms, such as k-means clustering, may be more appropriate. The k-means algorithm has different variations implemented in R, and it is important to choose the appropriate algorithm based on the specific requirements of the analysis.

The artificial data set used in this example is for a hypothetical mobile phone market and contains information on the number of features users want in a mobile phone and the price they are willing to pay for it. The data set is generated using the R programming language and the flexclust package.

The data set contains three distinct market segments, as shown in Figure 7.9. The bottom left segment represents users who want a cheap mobile phone with limited features, the middle segment represents users willing to pay slightly more for additional features, and the top right segment represents users willing to pay a higher price for advanced features.

To extract market segments from the data set, the k-means algorithm is used with the cclust() function from the flexclust package. The resulting segmentation solution is visualized using convex hulls, which are closed polygons connecting the outer points of each segment. Function clusterhulls() from the MSA package is used for this visualization.

It is noted that the segmentation solution may differ for different initial values in the k-means algorithm, as the algorithm starts with randomly selected initial segment representatives. Therefore, it is recommended to run the algorithm multiple times with different random initializations to eliminate any biased results.

To determine the optimal number of segments (clusters), the clustering procedure is repeated for different numbers of segments (from 2 to 8), and the sum of distances of all observations to their representative is compared across these solutions. A lower distance indicates a better segmentation solution as members of market segments are more similar to each other.

Finally, 10 runs of the k-means algorithm are calculated for each number of segments using different random initial representatives, and the best solution for each number of segments is retained for further analysis.

Algorithms with Integrated Variable Selection

Segmentation variables can impact the effectiveness of algorithms in extracting segments from data. Pre-processing methods can identify redundant or noisy variables. The filtering approach by Steinley and Brusco is introduced as an effective method for selecting segmentation variables, but it only works for metric variables. For binary data, selecting suitable segmentation variables can be more challenging, and the biclustering and VSBD algorithms are presented as potential solutions. Lastly, the approach of compressing segmentation variables into factors before segment extraction is discussed.

- Biclustering Algorithms - Biclustering algorithms exist for any kind of data, including metric and binary. Biclustering is particularly useful in market segmentation applications with many segmentation variables, because traditional clustering algorithms are not useful in this context due to the large numbers of genes which serve as variables for the grouping task. Biclustering experienced a big revival to address these challenges. There are several popular biclustering algorithms that exist; in particular, they differ in how a bicluster is defined. Biclustering is not one single very specific algorithm; rather it is a term describing a family of algorithms differing with respect to the properties of data they can accommodate.
- Variable Selection Procedure for Clustering Binary Data (VSBD) - The algorithm is based on the k-means clustering method and assumes the presence of masking variables. The method first identifies the best small subset of variables to extract segments, and then adds additional variables one by one, leading to the smallest increase in the within-cluster sum-of-squares criterion. The algorithm stops when the increase in within-cluster sum-of-squares reaches a threshold. The procedure recommends calculating the Ratkowsky and Lance index for the complete data with all variables to select the number of segments.
- Variable Reduction: Factor-Cluster Analysis - Factor-cluster analysis is a two-step procedure of data-driven market segmentation where the raw data, the original segmentation variables, are discarded, and the factor scores are used to extract market segments. This approach is conceptually legitimate only in a few cases, such as IQ tests. Simulations studies have shown that the number of consumers in a data set should be at least 100 times the number of segmentation variables. Factor analysis leads to a substantial loss of information and transforms the data. Therefore, the nature of the data is changed before segment extraction, and factor-cluster results are more difficult to interpret.

Data Structure Analysis

It introduces stability-based data structure analysis as an alternative approach to assessing the reliability and stability of segmentation solutions. The approach provides insights into the properties of data and can help identify natural, distinct, and well-separated market segments. Four different approaches to data structure analysis are discussed:

- **Cluster Indices** - The process of selecting the number of market segments to extract in market segmentation analysis is critical and requires guidance. Cluster indices are a common approach to obtain such guidance, and they can be classified into two groups: internal cluster indices and external cluster indices.

1. **Internal Cluster Indices** - Internal cluster indices can be used to determine the optimal number of segments in market segmentation. These indices measure the compactness and separation of each segment and require a distance measure between observations. The scree plot is a commonly used graph to select the number of segments based on the sum of within-cluster distances. Another index, the Ball-Hall index, corrects for the monotonous decrease of the internal cluster index with increasing numbers of market segments. Lastly, the Ratkowsky and Lance index uses the average value of observations within a segment to calculate an index for variable selection.
2. **External Cluster Indices** - External cluster indices are used to evaluate market segmentation solutions using additional external information, which could be the true segment structure of the data or the market segmentation solution obtained using a repeated calculation. To compare two segmentation solutions, external indices such as Jaccard and Rand indices are used, which focus on whether pairs of consumers are assigned to the same segments repeatedly rather than focusing on the segments' individual consumers are assigned to. However, the absolute values of these indices are difficult to interpret because minimum values depend on the size of the market segments contained in the solution, and hence the adjusted Rand index is used to solve this problem.

- **George Plots** – It is another method for assessing the separation of market segments, using the distance between consumers and segment representatives. The similarity of a consumer to the representative of a segment is calculated based on their distance, with a hyperparameter controlling the differences in similarity. Gorge plots can be used to visualize the similarity values, and the shape of the plot can indicate the presence of well-separated market segments.
- **Global Stability Analysis** - Resampling methods generate multiple segmentation solutions by creating new datasets using different algorithms, and their stability across repeated calculations is then compared to find the best segmentation solution. There are three categories of consumer data: distinct, well-separated market segments; unstructured data; and reproducible segmentation. The process of global stability analysis helps to determine which category of data set any given data falls under and the most appropriate number of segments to extract from the data.
- **Segment Level Stability Analysis** - Choosing the best overall segmentation solution may not necessarily mean it contains the most stable individual segment, so it's important to evaluate both global and individual segment stability when selecting a market segmentation solution.
    1. **Segment Level Stability Within Solutions** - Dolnicar and Leisch (2017) propose a new approach for assessing market segmentation solutions. Instead of evaluating the entire solution's stability, their criterion of segment level stability within solutions (SLSW) measures how often a market segment with the same characteristics is identified across a few repeated calculations of segmentation solutions with the same number of segments. The SLSW method calculates stability at the segment level and prevents an overall bad market segmentation solution containing one suitable market segment from being discarded.
    2. **Segment Level Stability Across Solutions** - SLSA helps to determine the re-occurrence of a market segment across different market segmentation solutions containing different numbers of segments. Natural segments are more attractive to organizations because they actually exist and no managerial judgment is needed in their artificial construction.

Step 6: **Profiling Segments(Mriganka)**

Identifying Key Characteristics of Market Segments

Profiling is necessary for data-driven market segmentation to identify defining characteristics of market segments. Good profiling is essential for correct interpretation of the resulting segments and making good strategic marketing decisions. Data-driven segmentation solutions are difficult to interpret, and graphical statistics approaches can make profiling less tedious and prone to misinterpretation.

Traditional Approaches to Profiling Market Segments

There are challenges of presenting data-driven segmentation solutions to clients. Exact percentages for each segment and segmentation variable are difficult to interpret and require extensive comparisons between segments and the total. This task becomes even more challenging when there are multiple segmentation solutions to compare. The use of statistical significance tests is not applicable in this case due to the nature of how segments are created.

Segment Profiling with Visualizations

The use of graphics is important in market segmentation analysis as they provide insights into complex relationships between variables and make results easier to interpret. Visualizations of segmentation solutions can help in inspecting and assessing the usefulness of a market segmentation solution and assist in making critical decisions in selecting one of the possible solutions.

- **Identifying Defining Characteristics of Market Segments -** Segment profile plots can be used to understand market segments. The plot is a visual representation of how each market segment differs from the overall sample for all segmentation variables. Segmentation variables can be ordered, marker variables can be defined to interpret the plot. A plot is also easier and faster to interpret than a table, and it can provide valuable insights for marketers.

- **Assessing Segment Separation -** Segment separation plots are used to depict the overlap of segments in all relevant dimensions of a data space. The plot consists of a scatter plot of observations coloured by segment membership and cluster hulls, and a neighborhood graph that indicates similarity between segments. The plot becomes complex as the number of segmentation variables increases, and therefore, projection techniques can be used to create a plot that maximizes separation. The advantages of the segment separation plot are combined with the advantages of perceptual maps to create an enhanced version. The plot becomes messy and hard to read when there is an overlap of market segments, and therefore, modifications like changing colors and highlighting the inner area of each segment can lead to a cleaner version.

Code: https://github.com/Mriganka-github/Market-Segmentation

Step 7: Describing Segments(Divyanshu)

Developing a Complete Picture of Market Segments

Understanding the variations in segmentation factors across market segments is the goal of segment profiling. Early in the market segmentation study process, segmentation variables are selected. The foundation for deriving market segments from empirical data is segmentation variables. In this phase, market segments are described using more data on segment participants. Crossing them with all other factors, such as media exposure, a particular product, brand attitudes or assessments, and psychographic, demographic, and socioeconomic data, can help further identify and characterize the segment. Gaining a thorough understanding of the characteristics of market segments depends on the quality of the segment descriptions. Segment descriptions are also necessary for creating a tailored marketing mix.

Using Visualisations to Describe Market Segments

Two significant benefits of using graphical statistics to describe market segments are that it makes it easier for users and data analysts to understand results and incorporates information on the statistical significance of differences, preventing the over-interpretation of insignificant differences.

Nominal and Ordinal Descriptor Variables

Cross-tabulating segment membership with the descriptor variable is the foundation for all visualizations and statistical tests when differentiating market segments using a nominal or ordinal descriptor variable. The cross-tabulation can be visualized with the help of stacked bar charts and mosaic charts. Additionally, mosaic plots may depict tables with more than two descriptor variables and incorporate inferential statistical components.

Metric Descriptor Variables

Using metric descriptor variables, conditional charts are a good tool for illustrating variations across market groups. Most common R plots may be found in conditional variants thanks to the R package lattice. The software ggplot2 offers a different approach for conditional charts. A parallel box-and-whisker plot, which displays the distribution of the variable individually for each segment, can be used to glean further insights. The value of a metric descriptor variable can be followed throughout several market segmentation solutions using a modified version of the segment-level stability across solutions (SLSA) display.

Testing for Segment Differences in Descriptor Variables

Formally testing for variations in descriptor variables across market groups may be done using straightforward statistical tests. Running separate tests for each relevant variable is the easiest method to look for differences. Segment membership is a nominal variable that may be handled like any other. It serves as the segmentation variables' nominal summary statistic. Therefore, any test to determine if a nominal variable is associated with another variable is appropriate. Cross-tabulation can be used to calculate the relationship between the nominal segment membership variable and another nominal or ordinal variable. The $\chi 2$-test is the proper test to determine whether columns and rows in a table are independent.

Parallel boxplots show the relationship between segment membership and metric variables. Any test for locational difference (mean, median) over different market segments can determine whether the reported locational differences are statistically significant. Analysis of Variance (ANOVA) is the most often used technique for determining if there are substantial differences between the means of more than two groups.

Predicting Segments from Descriptor Variables

A different strategy for understanding market segmentation is to predict segment membership based on descriptor variables. We employ a regression model in which the independent variables are descriptor variables, and the categorical dependent variable is segment membership. According to the descriptor variables, the prediction performance shows how effectively members of a market segment may be recognized.

Regression coefficients in linear regression models describe how much the dependent variable varies when one of the independent factors changes while the other independent variables stay the same. According to the linear regression model, changes brought about by adjustments to one independent variable are independent of changes in the absolute levels of the other independent variables. In the linear regression model, the dependent variable has a normal distribution. Generalized linear models may support a more extensive range of distributions for the dependent variable.

In generalized linear models, the dependent variable might be distributed according to the normal, Poisson, binomial, or multinomial distributions. For classification, a multinomial or binomial distribution is required. The generalized linear model is given by the following equation.

Binary Logistic Regression

By assuming that $f(y|\mu)$ is the Bernoulli distribution with success probability μ, and selecting the logit link that transfers the success probability $(0, 1)$ onto $(-\infty, \infty)$, we may create a regression model for binary data using generalised linear models.

When all of the independent variables (x1,..., xp) have values of 0, the intercept in a linear regression model calculates the mean value of the dependent variable. In a binomial logistic regression, if all of the independent variables (x1,..., xp) have values of 0, the intercept yields the value of the linear predictor η. When one independent variable change while others remain constant, the other regression coefficients in a linear regression model show how much the mean value of the dependent variable changes. The regression coefficients in binary logistic regression show how the linear predictor varies. Changes in the log probabilities of success match changes in the linear predictor. The ratio of the chances of success μ to the chances of failure 1−μ is known as the odds of success. Both success and failure are equally likely if the chances are 1. Success is more likely than failure if the chances are greater than 1.

Multinomial Logistic Regression

Multinomial logistic regression can fit a model that predicts each segment simultaneously. The dependent variable y is not binary because segment extraction often yields more than two market segments. Instead, it is categorical and is thought to adhere to a multinomial distribution, with the logistic function acting as the link function.

Tree-Based Methods

Regression and classifiers, a different modeling strategy, uses trees to predict a binary or categorical dependent variable from a set of independent factors. The benefits of classification and regression trees include their capacity to execute variable selection, straightforward integration of interaction effects, and simplicity of comprehension aided by visuals.

The model is fitted using a stepwise process in the tree method. Consumers are divided into groups based on one independent variable at each phase. The goal of the split is to produce groups that are as unmixed as feasible in terms of the dependent variable. This indicates that the dependent variable's values are comparable among customers in the resultant groupings. In the ideal scenario, a categorical dependent variable's value is the same for every group member.

The nodes that arise from each splitting phase are displayed in the resultant tree. The root node is the node that houses every consumer. Terminal nodes are nodes that cannot be further divided. Moving down the tree, we forecast section membership. We go along the branch that reflects the consumer's independent variable at each node. Based on the segment memberships of the consumers in the terminal node, we can forecast segment membership when we reach there.

Tree-building algorithms vary in terms of:

- splits that occur at each node into two or more groups (binary vs. multi-way splits).
- criteria for choosing the independent variable for the following split.

- Selection criteria for the independent variable's split point.
- Defining the stepwise process's stopping point.
- Final prediction at the terminal node

Checklist

- Bring over one or a few market segmentation options chosen based on compelling profiles from Step 6 (profiling).
- Select descriptor variables. Additional bits of data about every customer are known as descriptor variables and are used in market segmentation analysis
- Gain an understanding of the variations across market segments about descriptor variables by using visualization approaches. Use good plots, such as box-and-whisker plots for metric descriptor variables and mosaic plots for categorical and ordinal descriptor variables.
- Analyse the statistical impact of the descriptor variables.
- Correct for multiple testing if you conducted different statistical tests for each descriptor variable to prevent overestimating significance.
- To gauge your knowledge of each market area, "introduce" it to the other team members.
- Ascertain whether further information is needed to get a complete picture of particular portions.

Code: https://github.com/Divyan-shu-Singh/Market_Segmentation_macdonald-s/tree/main

Step 8: Selecting the Target Segment(s)(Divyanshu and Meenakshi)

The Targeting Decision

An effective marketing strategy tool is market segmentation. Choosing one or more target segments is a long-term choice that will significantly impact how well an organization performs in the future. Several segments are accessible for in-depth analysis when a worldwide market segmentation solution has been selected. Profiles and descriptions of these portions are provided. One or more of such market categories is necessary for targeting.

Questions that the segmentation team must answer can be divided into two categories:

- Which market category would the company like to focus on? Which market would the firm like to invest in?
- Which businesses selling the same thing would each group prefer to purchase from? How probable is it that our company will be selected? What is the likelihood that each part would agree to join us?

The article discusses the use of decision matrices to select a target market, with various versions of the matrix proposed in the past, including the Boston matrix, General Electric/McKinsey matrix, and McDonald’s four-box directional policy matrix. The aim is to evaluate alternative market segments and select one or a small number for targeting based on two criteria: segment attractiveness and relative organizational competitiveness. The article provides an example of a generic segment evaluation plot that can be produced in R and explains how to calculate the attractiveness value for each segment based on the weights assigned to segment attractiveness criteria and their values for each segment. The result is a weighted value for each segment's overall attractiveness, which is plotted on the x-axis of the segment evaluation plot. The article emphasizes the importance of returning to the specifications of an ideal target segment as identified in Step 2 of the market segmentation analysis, which resulted in a number of criteria of segment attractiveness and weights quantifying their impact. In Step 8, the target segment selection step, the actual value of each market segment for each attractiveness criterion is determined based on the grouping, profiling, and description of each segment. The location of each market segment on the segment evaluation plot is then computed by multiplying the weight of the segment attractiveness criterion with its value for each segment. The article also notes that there is no single best measure of segment attractiveness or relative organizational competitiveness, and it is up to the segmentation team to decide which variation of the decision matrix offers the most useful framework for decision-making. Overall, the article provides a clear and detailed guide for evaluating and selecting a target market using decision matrices and segment evaluation plots.

Step 9: Market Segment Evaluation

IMPLICATIONS OF MARKET MIX DECISION –

The article discusses the importance of market segmentation as a part of strategic marketing, along with its role in the segmentation-targeting-positioning approach. It highlights the need to customize the marketing mix to the target segment to maximize the benefits of a segmentation strategy, which can include changes to product, price, place, and promotion. The article also suggests that organizations can structure their market segmentation analysis around one of the 4Ps, depending on their specific marketing objectives. The paragraph discusses the implications of market segmentation for marketing mix decisions. The marketing mix was originally seen as a toolbox for achieving the best possible sales results, with marketers using various ingredients, such as product planning, pricing, advertising, and promotions, to achieve their objectives. Market segmentation is not an independent marketing strategy but rather goes hand in hand with other areas of strategic marketing, such as positioning and competition. The segmentation-targeting-positioning approach is a sequential process that starts with market segmentation, followed by targeting and positioning. However, it is important not to adhere too strictly to this process and to review each aspect of the marketing mix thoroughly once the target segment has been selected. To maximize the benefits of market segmentation, it is essential to customize the marketing mix to the target segment. One option is to structure the entire market segmentation analysis around one of the 4Ps. Overall, the article emphasizes the need for a thorough review of all aspects of the marketing mix once the target segment or segments have been selected.

PRODUCT

The product dimension of the marketing mix involves specifying a product based on customer needs and may involve modifying an existing product. Other marketing mix decisions under the product dimension include naming the product, packaging it, offering warranties, and after-sales support services. Product design or modification is driven by target segment selection, as illustrated by the market segments obtained from the Australian vacation activities data set using bi clustering. For instance, if targeting segment 3, which is interested in visiting museums, monuments, and gardens, a new product such as a MUSEUMS, MONUMENTS & MUCH, MUCH MORE product (accompanied by an activities pass) may be developed to help locate activities of interest to this segment. Alternatively, making gardens at the destination an attraction in their own right could also target this segment proactively.

PRICE

The article discusses the process of customizing the marketing mix for a destination wishing to market to a particular segment. The marketing mix involves several dimensions, including the price dimension, which requires decisions about setting prices and discounts for products. The article outlines the steps involved in creating a segment membership vector to compare members of a specific segment to tourists not belonging to that segment. It also discusses the creation of a binary variable indicating if a consumer is assigned to a particular segment or not. The article then discusses the use of a parallel boxplot to illustrate how the price dimension can be used to harvest the targeted marketing approach, indicating that members of the targeted segment have higher vacation expenditures per person per day than other tourists, suggesting that there is potential to attach a premium price to the product. The product dimension of the marketing mix involves specifying a product based on customer needs and may involve modifying an existing product. Other marketing mix decisions under the product dimension include naming the product, packaging it, offering warranties, and after-sales support services. Product design or modification is driven by target segment selection, as illustrated by the market segments obtained from the Australian vacation activities data set using biclustering. For instance, if targeting segment 3, which is interested in visiting museums, monuments, and gardens, a new product such as a MUSEUMS, MONUMENTS & MUCH, MUCH MORE product (accompanied by an activities pass) may be developed to help locate activities of interest to this segment. Alternatively, making gardens at the destination an attraction in their own right could also target this segment proactively.

PLACE

In marketing, the place dimension of the marketing mix refers to the distribution of the product to customers, including whether it should be sold online or offline, sold directly to customers or through intermediaries such as wholesalers or retailers. In the context of market segmentation, understanding the booking preferences of a particular segment can inform decisions about distribution channels. For example, if members of a certain segment tend to book hotels online, it would be important to offer an online booking option for hotels in that destination. The propBarchart function from the flexclust package can be used to visualize booking behavior for different segments, as shown in Figure 11.3. This information can be used to inform decisions about the distribution of other products, services, and activities for that segment. In addition to understanding the booking behavior of segment 3, it is also important to consider their preferences for purchasing other products and services related to cultural heritage. For example, do they prefer to purchase tickets to museums and monuments online or offline? This information can help the destination tailor their distribution strategy and ensure that the MUSEUMS, MONUMENTS & MUCH, MUCH MORE product is available through the preferred channels. Another important consideration in the place dimension is the use of intermediaries such as wholesalers and retailers. In some cases, it may be more efficient for the manufacturer to sell directly to customers, while in other cases, intermediaries may provide value in terms of reach and expertise. For example, a wholesaler may have established relationships with travel agents who can promote the destination and its products to potential customers. Similarly, a retailer may have a physical presence in a high-traffic area that can attract customers who may not have otherwise been aware of the product. It is also important to consider the level of control that the manufacturer wants to maintain over the distribution process. Selling directly to customers provides greater control over the product and the customer experience, but also requires more resources and expertise. Using intermediaries may provide greater reach and efficiency, but may result in a loss of control over the customer experience.

PROMOTION

In summary, the promotion component of the marketing mix involves developing an advertising message and identifying the most effective way to communicate this message to the target market. Other tools in this category include public relations, personal selling, and sponsorship. To determine the best information sources for reaching members of a specific segment, a comparison of the information sources used for their last vacation can be done using a propBarchart function in R. For example, members of segment 3 rely more on information provided by tourist centers and prefer TV channel 7. This information can be used to design the promotion component of the marketing mix by making specific information packs available in the local tourist information center and developing a media plan to target channel 7 viewers with the MUSEUMS, MONUMENTS & MUCH, MUCH MORE product.

In summary, the marketing mix consists of four components: product, price, place, and promotion. The decisions made in each of these components are critical to the success of the marketing strategy. In terms of the place dimension, it is important to consider how the product will be distributed to customers, whether it be online or offline, direct to customers or through intermediaries such as wholesalers or retailers. Understanding the booking preferences of different market segments can help inform these decisions. Similarly, effective promotion requires a clear message that resonates with the target market, and identifying the most effective communication channels. In the case of segment 3, for example, the preference for information provided by tourist centres and TV channel 7 can inform the development of a media plan that maximizes exposure to the targeted communication of the MUSEUMS, MONUMENTS & MUCH, MUCH MORE product.

https://github.com/Meenu2204/Feynnlabs-Market-segmentation-/commit/56f33fec04927f7aed740d0f30f5c2f3e8b544a8