MARKET SEGMENTATION

Definitions of Market Segmentation:

Market segmentation refers to the process of dividing a large market into smaller groups of consumers or businesses with similar needs, characteristics, or behaviors. These groups are known as market segments, and they can be defined by various criteria, such as demographics (age, gender, income), psychographics (values, beliefs, attitudes), geography (region, city), or behavior (buying habits, usage rate).

Market segmentation helps businesses to understand their customers better and tailor their marketing strategies and products to meet the specific needs and preferences of each segment. By targeting specific segments, businesses can increase their effectiveness and efficiency in reaching and satisfying their customers, leading to increased sales and profits.

The Benefits of Market Segmentation:

Market segmentation is a useful tool for organizations to reflect on their strengths and gain insights into consumer needs. It leads to a better understanding of consumer differences and a better match between organizational strengths and consumer needs, resulting in a long-term competitive advantage. Market segmentation can be taken to the extreme by offering customized products to small groups of consumers. It also leads to a higher return on investment as marketing efforts are focused on consumers whose needs can be satisfied. Market segmentation can also contribute to sales management and team building by requiring representatives from different organizational units to work together.

The cost of Market Segmentation:

Implementing market segmentation requires a significant investment of time, money, and resources. The process involves a thorough analysis and development of a customized marketing mix. Continuous monitoring is also required to evaluate the success and modify the strategy accordingly. If implemented poorly, it can lead to substantial expenses with no return, which may disenfranchise staff involved in the exercise

The layers of market segmentation analysis:

1. Making it happen in practice:

Deciding to segment, defining the ideal segment, selecting (the) target segment(s),

developing a customised marketing mix, assessing effectiveness and monitoring

marketing changes

2. Enabling high quality market segmentation analysis:

Collecting good data, exploring data, profiling segments, describing segments

3. Conducting high quality market segmentation analysis:

Extracting market segments

Steps of Market Segmentation Analysis

Step 1: Deciding (not) to Segment:

Implications of Committing to Market Segmentation and its barriers:

Market segmentation is a key marketing strategy used by organizations, but it is not always the best decision to pursue such a strategy. Before investing time and resources in a market segmentation analysis, it is important to understand the implications of pursuing a market segmentation strategy. The key implication is that the organization needs to commit to the segmentation strategy on the long term. 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 organizational 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 organizational levels and across all organizational units.

Barriers that can impede the successful roll-out of a market segmentation strategy include lack of leadership, pro-active championing, commitment and involvement in the market segmentation process by senior leadership, lack of resources, lack of market or consumer orientation, resistance to change and new ideas, lack of creative thinking, bad communication and lack of sharing of information and insights across organizational units, short-term thinking, unwillingness to make changes, office politics, lack of training, lack of a qualified marketing expert in the organization, lack of a qualified data manager and analyst in the organization, and objective restrictions faced by the organization.

Step 2: Specifying the Ideal Target Segment:

Segment Evaluation Criteria:

The organisation has to determine two sets of segment evaluation criteria. The first set is the knock-out criteria, and the second set is the attractiveness criteria.

1.The knock-out criteria

The knock-out criteria are essential as they help to eliminate the segments that do not meet the minimum requirements of the organisation. The knock-out criteria can be different for different organisations, and they can change over time. For example, a knock-out criterion for a luxury car manufacturer may be that the segment must have a high income level, whereas a knock-out criterion for a fast-food chain may be that the segment must be looking for a quick and affordable meal.

Some examples of knock-out criteria proposed in the literature include:

- Measurable: The segment can be measured in terms of size and profitability.
- Substantial: The segment is large enough to be profitable.
- Accessible: The segment can be reached and served effectively.
- Differentiable: The segment is distinct from other segments.
- Homogeneous within: The segment members have similar needs and preferences.
- Heterogeneous between: The segment is different from other segments in terms of needs and preferences.
- Compatible with the company: The segment is compatible with the company's strengths and image.
- Actionable: The organisation can design and implement a marketing mix that appeals to the segment.

2. the attractiveness criteria

The criteria are not binary and each market segment is rated based on its level of attractiveness across various criteria. The overall attractiveness of a segment is determined by its performance across all criteria.

3. Implementing a Structured Process

The implementation of a structured process for market segmentation analysis involves determining the segment attractiveness and organizational competitiveness values, which are determined by the segmentation team. A large number of possible criteria need to be investigated before agreement is reached on which criteria are most important for the organization, and a team of people should be involved in this process. Representatives from a wide range of organizational units should be included in this process because each unit has a different perspective on the business of the organization, and the segmentation strategy will affect every single unit of the organization. The market segmentation team should have a list of approximately six segment attractiveness criteria, each with a weight attached to it to indicate how important it is to the organization compared to the other criteria. The typical approach to weighting is to ask all team members to distribute 100 points across the segmentation criteria, and these allocations have to be negotiated until agreement is reached.

Step 3: Collecting Data

Segmentation Variables:

Segmentation variables are the variables in the empirical data that are used to split a sample into market segments in market segmentation. In commonsense segmentation, the segmentation variable is typically one single characteristic of the consumers in the sample. For example, in a sample of consumers, gender could be used as a segmentation variable to create two segments: one segment of women and another of men. In data-driven market segmentation, however, multiple segmentation variables are used to create market segments. These variables serve as the starting point for identifying naturally existing, or artificially creating market segments useful to the organization. For example, a set of benefits sought when going on vacation could be used as segmentation variables to create segments of tourists with common preferences, regardless of gender. The other variables in the data, such as socio-demographics, travel behavior, and media behavior, are then used as descriptor variables to describe these segments in detail.

Segmentation Criteria:

The organization must choose a segmentation criterion before extracting segments and collecting data for segmentation. The segmentation criterion refers to the nature of the information used for market segmentation, and the most common criteria are geographic, socio-demographic, psychographic, and behavioral. There are various segmentation criteria available, and it is recommended to use the simplest approach that works for the product or service at the least possible cost. Factors such as profitability, bargaining power, preferences for benefits or products, barriers to choice, and consumer interaction effects are relevant in terms of market segmentation. The decision of which segmentation criterion to use requires prior knowledge about the market and cannot be easily outsourced to a consultant or data analyst.

1. Geographic Segmentation

Geographic segmentation is the original and simplest criterion for market segmentation. It assigns consumers to segments based on their location of residence. The advantage is that it is easy to target communication messages and select communication channels for each segment. However, it may not always be the best approach since people living in the same area may not necessarily share other relevant characteristics. International market segmentation studies have revived the use of geographic information as a segmentation variable, but it requires meaningful variables across all included geographic regions and may suffer from biases if completed by respondents from different cultural backgrounds. Examples of companies using geographic segmentation include Amazon and IKEA.

2. Socio-Demographic Segmentation

The socio-demographic segmentation criteria include age, gender, income, and education, which can be useful in industries such as luxury goods, cosmetics, baby products, retirement villages, and tourism resort products. However, segment membership based on socio-demographic criteria does not always provide sufficient market insight for optimal segmentation decisions, as demographics only explain a small percentage of the variance in consumer behavior. Values, tastes, and preferences are more influential in terms of consumers' buying decisions, according to Yankelovich and Meer.

3. Psychographic Segmentation

Psychographic segmentation groups people based on psychological criteria such as beliefs, interests, preferences, aspirations, or benefits sought when purchasing a product. Benefit segmentation and lifestyle segmentation are popular approaches to psychographic segmentation. Psychographic criteria are more complex than geographic or socio-demographic criteria, and it is difficult to find a single characteristic of a person that will provide insight into the psychographic dimension of interest. Therefore, most psychographic segmentation studies use a number of segmentation variables. Psychographic segmentation provides insight into the underlying reasons for differences in consumer behavior, but determining segment memberships for consumers is more complex. The power of the psychographic approach depends heavily on the reliability and validity of the empirical measures used to capture the psychographic dimensions of interest.

4. Behavioural Segmentation

Behavioural segmentation involves grouping consumers based on their actual or reported behavior, such as prior experience with a product, purchase frequency, amount spent on purchases, and information search behavior. This approach is advantageous because it uses the behavior of interest as the basis for segment extraction. However, behavioral data may not always be readily available, especially when studying potential customers who have not previously purchased the product.

Data from Survey Studies

1. Choice of Variables

The choice of variables included in market segmentation analysis is critical to the quality of the solution. In data-driven segmentation, all relevant variables should be included while avoiding unnecessary ones that can cause respondent fatigue and make the segmentation problem unnecessarily difficult. Noisy variables, which do not contribute to identifying the correct market segments and can negatively impact the algorithm's ability to extract the correct solution, should be avoided. Careful questionnaire development and variable selection, based on exploratory or qualitative research, can help ensure that no important variables are omitted while avoiding redundancies that can interfere with the segmentation analysis.

2. Response Options

The response options provided in surveys can affect the type of data available for subsequent analysis. Binary or dichotomous responses can be represented by 0s and 1s, while nominal variables can be transformed into binary data. Metric data, which allow any statistical procedure to be performed, are well suited for segmentation analysis. Ordinal data, generated by limited, ordered response options, pose difficulties for standard distance measures. Binary or metric response options are preferable to prevent complications in segmentation analysis. Visual analogue scales can be used to capture fine nuances of responses. Binary response options have been shown to outperform ordinal options in many contexts.

3. Response Styles

The response styles of survey participants can impact the accuracy of market segmentation analysis. Response bias, or the tendency to respond on a basis other than the specific item content, can result in misleading data. Common response styles include extreme answers, midpoints, and agreeing with all statements. If not accounted for, response styles can lead to misinterpretation of market segments and must be minimized or removed before conducting analysis.

4. Sample Size

The sample size is important in market segmentation analysis because it can significantly affect the accuracy and reliability of the segmentation results. Insufficient sample size can make it impossible to determine the correct number and nature of segments in the data set. On the other hand, increasing the sample size can improve the correctness of the extracted segments, but the marginal benefit of further increasing the sample size decreases as the sample size gets larger.

Data from Internal Sources

Internal data, such as scanner data, booking data, and online purchase data, represent actual behavior of consumers and are automatically generated, making them easy to access. The advantages of using internal data include the absence of response biases and the lack of effort required to collect the data. However, the danger of using internal data is that it may be systematically biased by over-representing existing customers and may not provide information about other consumers the organization may want to win as customers in the future.

Data from Experimental Studies

Experimental studies is another potential source of data for market segmentation analysis. Such studies may involve field or laboratory experiments, choice experiments, or conjoint analyses. These studies aim to test how consumers respond to specific product attributes and can provide valuable information about consumer preferences that can be used as segmentation criteria.

Step 7: Describing Segments

Developing a Complete Picture of Market Segments

The process of market segmentation involves selecting segmentation variables, extracting market segments from data, and describing segments using additional information or descriptor variables. Profiling involves understanding differences in segmentation variables across market segments, while segment description involves using additional information to describe segments. Good descriptions of market segments are crucial for developing a customized marketing mix. Differences between market segments with respect to descriptor variables can be studied using descriptive statistics and visualizations or inferential statistics.

Using Visualisations to Describe Market Segments

The use of visualisations to describe differences in descriptor variables between market segments. Visualisations are useful for simplifying interpretation for both data analysts and users, and integrating information on statistical significance to avoid over-interpretation of insignificant differences. Two basic approaches for visualisation are discussed for nominal/ordinal variables and metric variables. The use of graphical statistics is considered to be the essence of marketing research and is preferred by marketing managers due to its intuitiveness and efficiency compared to tabular results.

1. Nominal and Ordinal Descriptor Variables

In market segmentation analysis, the goal is to identify distinct groups of customers or consumers with similar needs, preferences, behaviors, or characteristics. Once these segments are identified, it is important to describe them in a meaningful way to gain insights and inform marketing strategies.

One way to describe segments is by using descriptor variables, which are characteristics that distinguish the segments from each other. These descriptor variables can be nominal (categorical) or ordinal (ordered categorical). Examples of nominal descriptor variables include gender, race, and geographic location, while examples of ordinal descriptor variables include income level, education level, and age group.

To describe segments using descriptor variables, a cross-tabulation of segment membership with the descriptor variable is typically used as the basis for visualizations and statistical tests. This allows for easy comparison of the differences between segments in terms of the descriptor variable.

Visualizations such as stacked bar charts and mosaic plots can be used to visualize the cross-tabulations and highlight differences between segments. In addition, inferential statistics such as standard deviations, Pearson residuals, and expected and observed frequencies can be used to assess the significance of differences between segments.

Overall, describing segments in market segmentation analysis is an important step in understanding customer needs and preferences and informing marketing strategies. By using descriptor variables and visualizations, marketers can gain valuable insights into the differences between segments and tailor their marketing efforts to better meet the needs of each segment.

Testing for Segment Differences in Descriptor Variables

To test for segment differences in descriptor variables, you can use statistical methods such as analysis of variance (ANOVA) or t-tests.

ANOVA is used to compare means across two or more groups. It tests the null hypothesis that the means of all groups are equal. If the p-value of the ANOVA test is less than a predetermined significance level (usually 0.05), then we reject the null hypothesis and conclude that at least one group is different from the others.

If you want to compare the means of just two groups, you can use a t-test. A t-test tests the null hypothesis that the means of the two groups are equal. If the p-value of the t-test is less than the significance level, then we reject the null hypothesis and conclude that the means of the two groups are different.

When conducting these tests, it's important to consider the assumptions of the statistical method being used. For ANOVA and t-tests, the main assumptions are normality and equal variances. Normality assumes that the data is normally distributed, and equal variances assumes that the variance of each group is equal. If these assumptions are not met, you may need to use a different statistical method or consider data transformation techniques.

Predicting Segments from Descriptor Variables

In marketing research, segmenting the market is crucial to identify target groups of customers with similar needs and preferences. There are several ways to segment the market, including cluster analysis and factor analysis. Another way is to use regression analysis to predict segment membership from descriptor variables. Regression models assume that a dependent variable can be predicted from independent variables or regressors. In market segmentation, the dependent variable is the segment membership, which is a categorical variable, and the independent variables are the descriptor variables.

Linear regression models assume that the relationship between the dependent variable and the independent variables is linear, and the dependent variable follows a normal distribution. The regression coefficients in linear regression models express how much the dependent variable changes when one independent variable changes, while all other independent variables remain constant. The regression coefficients also indicate the mean difference in the dependent variable between different categories of the categorical variable.

Generalized linear models can accommodate a wider range of distributions for the dependent variable, including categorical variables. In this case, the mean value of the dependent variable can only take values between 0 and 1. The link function in generalized linear models transforms the mean value to a scale where it can take any real value, and the regression coefficients still express the change in the dependent variable caused by a change in an independent variable.

Predicting segment membership from descriptor variables using regression models helps identify which descriptor variables are critical to the identification of segment membership. The prediction performance indicates how well members of a market segment can be identified given the descriptor variables.

1. Binary Logistic Regression

Binary logistic regression is a commonly used method in market segmentation analysis to predict the likelihood of consumers belonging to a specific segment based on their demographic and psychographic characteristics. In this approach, the dependent variable is a binary indicator of whether a consumer belongs to a specific segment, while the independent variables are demographic and psychographic characteristics such as age, gender, income, education level, attitudes, values, and behaviors.

To perform binary logistic regression in market segmentation analysis using Python, one can use the logistic regression module from the scikit-learn library. The logistic regression module implements logistic regression using a generalized linear model framework. The logistic regression model is specified with the "logistic" argument for the solver parameter, which uses a logit link function for the binary response variable.

create a logistic regression object

logreg = LogisticRegression(solver='liblinear', random_state=0)

fit the model using training data

logreg.fit(X_train, y_train)

make predictions using test data

y_pred = logreg.predict(X_test)

The output of the logistic regression model includes regression coefficients for each independent variable, the intercept, and information on the model fit such as degrees of freedom, null deviance, residual deviance, and AIC. The intercept represents the log odds of success (i.e., belonging to the segment of interest) when all independent variables are equal to zero. The regression coefficients represent how the log odds of success change when the independent variables change.

To interpret the regression coefficients and their effects in Python, one can use the effects package, which is a Python implementation of the R effects package. The effectplot() function in the effects package calculates the predicted values for different levels of the independent variables while keeping other independent variables constant at their average value. The predicted values are the probabilities of belonging to the segment of interest. The results can be plotted to visualize how the predicted probabilities change with different independent variables.

Create an effect object

eff = Effect(model, default_levels = {'independent_variable_1': [0, 1], 'independent_variable_2': [0, 1]})

Overall, binary logistic regression is a useful tool for market segmentation analysis as it can help identify the demographic and psychographic characteristics of consumers who are more likely to belong to a specific segment. This information can be used to tailor marketing strategies and product offerings to better meet the needs and preferences of the target segment.

2. Multinomial Logistic Regression

Multinomial logistic regression is a statistical technique used to model and analyze data with categorical outcomes that have more than two categories. It is a useful technique for market segmentation, which is the process of dividing a market into groups of consumers with similar needs or behaviors.

In market segmentation, multinomial logistic regression can be used to identify the characteristics of different consumer segments and understand how those characteristics influence consumer behavior. For example, a company might use multinomial logistic regression to identify different consumer segments based on demographic variables such as age, gender, and income, as well as behavioral variables such as purchasing habits and product preferences.

To perform multinomial logistic regression, you need a dataset with a categorical dependent variable that has more than two categories and one or more independent variables that may influence the dependent variable. The dependent variable represents the market segments you want to analyze, and the independent variables represent the characteristics or behaviors that may differentiate those segments.

In Python, you can perform multinomial logistic regression using the `LogisticRegression` class from the scikit-learn package. The class takes input data and target labels as input, and returns a fitted model object that can be used to make predictions and analyze the relationships between the dependent and independent variables.

```
# Fit the model
model = LogisticRegression(multi_class='multinomial', solver='lbfgs')
model.fit(X, y)
```

Once you have fitted the model, you can use the coefficients to understand the effects of each independent variable on the probability of belonging to each market segment. You can also use techniques such as cross-validation and model selection to evaluate the performance of the model and improve its accuracy.

Overall, multinomial logistic regression is a powerful tool for market segmentation that can help businesses identify and target different consumer segments more effectively. By understanding the characteristics and behaviors of different segments, companies can tailor their marketing strategies and product offerings to better meet the needs of their customers and increase their competitiveness in the marketplace.

3. Tree-Based Methods

Tree-based methods are a type of supervised learning technique used for predicting binary or categorical dependent variables based on independent variables. The classification and regression trees (CARTs) are an example of tree-based methods, and they have advantages such as performing variable selection, easy interpretation, and straightforward incorporation of interaction effects. However, the results of tree-based methods can be unstable due to small changes in the data.

The tree-based approach involves a stepwise procedure that splits consumers into groups based on one independent variable to make the resulting groups as pure as possible with respect to the dependent variable. The resulting tree shows the nodes that emerge from each splitting step, and terminal nodes are not split any further. We predict segment membership by moving down the tree and predict the membership based on the segment memberships of consumers contained in the terminal node.

Tree constructing algorithms differ in several ways, such as the splits into two or more groups at each node, the selection criterion for the independent variable for the next split, the selection criterion for the split point of the independent variable, the stopping criterion for the stepwise procedure, and the final prediction at the terminal node.

There are several Python packages that implement tree constructing algorithms, such as scikit-learn and XGBoost. The scikit-learn package implements decision trees and random forests, while the XGBoost package implements gradient boosting trees. Additionally, the dtreeviz package provides a visualization tool for decision trees.

An alternative tree constructing procedure that performs unbiased variable selection and allows visualization of the fitted tree models is the conditional inference tree, which can be implemented using the implementation provided in the partykit Python package. We can use the ctree() function from the partykit package to fit a conditional inference tree and predict segment membership in a data set.