**Summarization of Market Segmentation Analysis**

**Adwait Gore**

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

1. **Implications of committing to market segmentation**

* Long-term organizational commitment and willingness to make substantial changes
* Investment of money in market research, focus groups, field surveys and advertisements.
* Potentially required changes include change in pricing, new product development and modification of existing products, etc.
* To maximize the benefits focus should be on market segments rather than products.
* Segmentation strategy decision must be made at highest level and continuous communication with all organizational levels and across all units.

1. **Implementation barriers**

* Lack of leadership and involvement of senior management in segmentation process.
* Lack of creative thinking and resistance to change.
* Lack of financial resources and inability to make structural changes.
* Lack of qualified marketing experts in the organizations.
* Lack of planning or bad planning to exercise the market segmentation and having no clear objectives about it.

**Adwait Gore**

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

1. **Segment evaluation criteria**

* This is a third layer of market segmentation which depends on user inputs.
* User input should be throughout the process of market segmentation analysis.

1. **Knock-out criteria**

It is used to determine if market segments identified in segmentation analysis qualify for assessment using segment attractiveness criteria.

* The segment must be homogenous.
* The segment must be distinct.
* The segment must be large enough and matching the strengths of the organization
* Members of the segment must be identifiable (possible to spot them in the marketplace).
* Segment must be reachable.

The senior management, the segmentation team, and the advisory committee must have clear idea about knock-out criteria.

1. **Attractiveness criteria**

* Attractiveness criteria are not binary in nature.
* Each market is rated (more attractive or less attractive) with respect to a specific criterion.
* These attractiveness criteria determines whether a market segment is selected as a target segment in market segmentation analysis.

1. **Implementing a structured process**

* Most popular structured approach for evaluating market segments is the use of segment evaluation plot.
* A team of two to three people is sufficient who can propose their initial solution to advisory committee which consists of representatives of all organizational units for discussion and possible modification.
* Approximately six segment attractiveness criteria have to be listed and weighted accordingly by distributing 100 points across the segmentation criteria by all team members to compare with other criteria.

**Adwait Gore**

**Step 3: -** **Collecting data**

1. **Segmentation variables**

* Empirical data forms the basis of both commonsense and data-driven market segmentation.
* It is important to identify or create market segments and also in developing a valid segmentation solution.
* In commonsense segmentation, the segmentation variable is typically one single characteristic of that consumer, while data-driven segmentation relies on multiple variables rather than just one.
* Empirical data can come from wide range of sources such as, survey studies, scanner data, experimental studies, etc.

1. **Segmentation criteria**

The term “Segmentation variable” refers to one measured value but “Segmentation Criterion” relates to the nature of the information used for market segmentation.

1. **Geographic segmentation: -** In this, the consumer’s location of residence used as the only criterion to form market segments.
2. **Socio-Demographic segmentation: -** It includes age, gender, income and education. It can be very useful in some industries.
3. **Psychographic segmentation: -** When people are grouped according to psychological criteria, such as their beliefs, interests, preferences, aspirations, or benefits sought when purchasing a product.
4. **Behavioural: -** In this method, customer’s prior experience with the product, frequency of purchase, amount spent on purchasing the product on each occasion. This kind of information is used for market segmentation.
5. **Data from survey studies**
6. **Choice of Variables: -** Carefulselection of variables is necessary for the better performance of the algorithm. The unnecessary variables termed as “Noisy variables” which do not contribute any information necessary for the identification correct market segments can negatively impact the segmentation solution if included.
7. **Response Options: -** Binary and metric data are preferable as they facilitate distance measurement essential for market segmentation.
8. **Response Styles: -** Common responses styles include tendencies to choose extreme options, the midpoint, or to agree with all statements which can distort segmentation results.
9. **Sample Size: -** Adequate sample size is required for statistical validity. Too small sample data may not give the clear picture and too large data can lead to unnecessary computation cost and data complexity**.**
10. **Data from Internal Sources**

* Organizations have substantial internal data for market segmentation such as scanner data, online purchase data, etc.
* This data represents actual consumer behaviour.
* A potential drawback is systematic bias by over representing existing customers and lacks information about potential future customers who may have different consumption patterns.

1. **Data from experimental Studies: -** Experimental data can result from field or laboratory experiments. Experimental data can also result from choice experiments or conjoint analyses. The aim of such studies is to present consumers with carefully developed stimuli consisting of specific levels of specific product attributes.

**Adwait Gore**

**Step 4: -** **Exploring Data**

1. **A first glimpse at the data**

* The Exploratory data analysis and data cleaning and pre-processing are important to analyse the trends in the data.
* To understand the trends, we use python libraries like pandas for data cleaning purpose and matplotlib, seaborn for visualization purpose.

1. **Data Cleaning**

* The first step before doing data analysis is to clean the given data.
* This includes check for null values, duplicate values, using measures of central measures (Mean, Mode, Median) to replace null values, etc.
* Visualize the data to check the outliers [using python command df. plot (kind = box)].

1. **Descriptive Analysis**

* Descriptive numeric and graphic representation provide insights into the data.
* To obtain numeric summary in python use command df. describe (). This gives you the five number summary of data which includes [max, min, Q1(25 percentile), Q2=Median (50% percentile), Q3 (75 percentile), count, standard deviation].
* Histogram, boxplot and scatter plot of numeric data helps to visualize the data and to understand the trends in the data.
* Histogram shows how often observations within a certain value range occur. It also shows that whether data is skewed or normally distributed.
* Boxplot gives the outliers present in the data.

1. **Pre-processing:**
2. **Categorical Variables**

* Two pre-processing procedures are often used for categorical variables.

1) merging of categorical levels before further analysis.

2) converting the categorical variables to numeric one. This is done by two methods I) using label encoder II) one hot encoding

**3)** These encoded columns are now used for further analysis.

1. **Numeric Variables**

* The range of values of a segmentation variable affects its relative influence in distance-based methods of segment extraction.
* To balance the influence of segmentation variables on segmentation results, variables can be standardized.
* Standardizing variables means transforming them in a way that puts them on a common scale.
* In python’s sci-kit learn library there are two scalers 1) MinMaxScaler 2) StandardScaler
* MinMaxScaler converts the data between minimum and maximum value of that particular column that needs to be standardized. On the other hand, StandardScaler converts that numerical column data between 0 and 1.

1. **Principal Component Analysis**

* It transforms the multivariate data set containing metric variables to a new data set with new variables referred to as principal components.
* These principal components are uncorrelated and ordered by importance.
* Principal components analysis basically keeps the data space unchanged, but looks at it from a different angle.
* Works off the covariance or correlation matrix of several numeric variables.
* If the data ranges are different, the correlation matrix should be used (equivalent to standardizing the data)
* PCA typically used to project high-dimensional data into lower dimension for plotting purposes.

**Adwait Gore**

**Step 7: -** **Describing Segments**

1. **Developing a complete picture of Market Segments**

* Segment profiling involves understanding the differences in in segmentation variables across market segments.
* Descriptive statistics, visualization, inferential statistics are some methods to understand these differences.

1. **Using visualizations to describe market segments**

* A wide range of charts are available to visualize the descriptor variable differences.
* Visualization simplifies the interpretation for both data analyst and the user.
* People process graphical data easily as compared to tabular data.

1. **Nominal and Ordinal Descriptor Variables**

* Use of descriptor variables are required to describe the market and effective segmentation.
* The use of segment plot and segment number will effectively project the different variables that are influencing the segment.

1. **Metric Descriptor Variables**

* Use of metric variables with continuous values and trends of the variables is key to segmentation and its success.
* Conditional plots are well-suited for visualising differences between market segments using metric descriptor variables.

1. **Testing for Segment Differences in Descriptor Variables**

The use of statistical methods and tests are required for checking the success of the segmentation. The testing of model success using f1 score and R2 score will also be useful for successful segmentation**.**

1. **Predicting Segments from Descriptor Variables**

* To learn market segments, regression model is used with segment variables as categorical dependent variable and descriptor variables as dependent variables.
* This is done by using supervised machine learning models.
* Regression analysis is the basis of model prediction.

1. **Binary Logistic Regression**

* This model is used when data has binary descriptor variables.

1. **Multinomial Logistic Regression**

* Multinomial logistic regression can fit a model that predicts each segment simultaneously. Because segment extraction typically results in more than two market segments, the dependent variable y is not binary.
* The regression coefficients are arranged in matrix form.

1. **Tree Based Methods**

* Classification and regression trees are a supervised learning technique from machine learning.
* The advantages of classification and regression trees are their ability to perform variable selection, ease of interpretation supported by visualizations, and the straight-forward incorporation of interaction effects.
* The tree approach uses a stepwise procedure to fit the model.
* Because of this stepwise splitting procedure, the classification and regression tree approach are also referred to as recursive partitioning.