Feynn Labs Report -Daksh Agiwal

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

The text discusses the implications and potential barriers of implementing a market segmentation strategy. Market segmentation requires a long-term organizational commitment and substantial changes, such as developing new products, modifying existing products, changing pricing and distribution channels, and adjusting the internal structure of the organization. The decision to investigate the potential of a market segmentation strategy must be made at the highest executive level and systematically and continuously communicated and reinforced at all organizational levels. The potential barriers to implementing market segmentation include lack of leadership, commitment, involvement, and resources from senior management, organizational culture, lack of training, lack of formal marketing function, and objective restrictions faced by the organization. The text emphasizes that the successful implementation of market segmentation requires pro-active leadership, market orientation, creative thinking, good communication.

**Step 2: Specifying the Ideal Target Segment**

The third layer of market segmentation analysis is based on user input, and the organisation must make a major contribution to market segmentationanalysis in Step 2. This includes determining two sets of evaluation criteria: knock-out criteria, which are essential, non-negotiable features of segments that the organisation would consider targeting, and attractiveness criteria, which evaluate the relative attractiveness of the remaining market segments. The literature proposes a wide array of possible segment evaluation criteria, and Table 4.1 contains a selection of proposed criteria. It is not up to the segmentation team to negotiate the extent to which they matter in target segment selection. The second set of attractiveness criteria is a shopping list for the segmentation team to select which of these criteria they want to use to determine how attractive potential target segments are. Knock-out criteria are used to determine if market segments resulting from the market segmentation analysis qualify to be assessed using segment attractiveness criteria. The ideal target segment must be large enough, homogeneous, distinct, identifiable, reachable, and matching the strengths of the organisation.

**Step 3: Collecting Data**

Empirical data is used to identify or create market segments and describe these segments in detail. In commonsense segmentation, the segmentation variable is typically one single characteristic of the consumers in the sample. In data-driven market segmentation, multiple segmentation variables are used to identify naturally existing, or artificially creating market segments useful to the organisation. Examples include socio-demographics, media behaviour, and demographic characteristics. Empirical data for segmentation studies can come from a range of sources, such as survey studies, scanner data, loyalty programs, and experimental studies. Data quality is critical to both assigning each person in the sample to the correct market segment, and being able to correctly describe the segments. Good market segmentation analysis requires good empirical data. The organisation must make an important decision about which segmentation criterion to use for market segmentation, which can be outsourced to a consultant or data analyst. The most common segmentation criteria are geographic, socio-demographic, psychographic and behavioural. Cahill (2006) states that the simplest approach is to use the least you can. Geographic segmentation is the most appropriate approach for market segmentation, as it allows consumers to be assigned to a geographic unit. However, living in the same country or area does not necessarily mean that people share other characteristics relevant to marketers, such as benefits they seek when purchasing a product. For example, people living in luxury suburbs may not be a good target market for luxury cars. Socio-demographic segmentation criteria can be useful in some industries, such as luxury goods, cosmetics, baby products, retirement villages, and tourism resort products. However, they may not provide sufficient market insight for optimal segmentation decisions. Psychographic segmentation is when people are grouped according to psychological criteria, such as beliefs, interests, preferences, aspirations, or benefits sought when purchasing a product. Psychographic segmentation is an umbrella term used to cover all measures of the mind. Benefit segmentation, which Haley (1968) is credited for, is the most popular kind of psychographic segmentation. Lifestyle segmentation is based on people's activities, opinions and interests. Psychographic criteria are more complex than geographic or socio-demographic criteria because it is difficult to find a single characteristic of a person that will provide insight into the psychographic dimension of interest. Behavioural segmentation is another approach to segment extraction that looks for similarities in behaviour or reported behaviour. In a comparison of different segmentation criteria used as segmentation variables, behaviours reported by tourists emerged as superior to geographic variables. Market segmentation analyses are based on survey data, which can be susceptible to a range of biases that can negatively affect the quality of solutions. Choosing the variables that are included as segmentation variables in commonsense segmentation and data-driven segmentation is critical to the quality of the solution. Unnecessary variables can make questionnaires long and tedious, leading to respondent fatigue and lower quality responses. The appropriate ratio of variables and the available sample is discussed later in this chapter. Noisy variables or masking variables can prevent algorithms from identifying the correct market segmentation solution. To avoid this, survey research should ask all necessary and unique questions, resist the temptation to include unnecessary or redundant questions, and develop a good questionnaire using exploratory or qualitative research. Response options should also be carefully selected to ensure that no critically important variables are omitted. Survey data is prone to capturing biases, such as a response bias, which is a systematic tendency to respond to a range of questionnaire items on some basis other than the specific item content. Response styles affect segmentation results, and it is important to minimise the risk of capturing response styles when data is collected for market segmentation. The importance of sufficient sample size in market segmentation analysis is highlighted by Figure 5.1, which illustrates the problem any segmentation algorithm faces if the sample is insufficient. Viennese psychologist Formann (1984) recommends that the sample size should be at least 2p (better five times 2p). Qiu and Joe (2015) developed a sample size recommendation for constructing artificial data sets for studying the performance of clustering algorithms. Experimental data can be used to form the basis of market segmentation analysis, such as field or laboratory experiments, choice experiments, and conjoint analyses. These studies present consumers with specific levels of product attributes and ask them to indicate which of the products they prefer. This information can be used as a segmentation criterion.

**Step 4: Exploring Data**

Data cleaning, also known as data preprocessing, is an essential step in machine learning. It refers to the process of identifying and correcting errors, inconsistencies, and missing data in the dataset before it is used for analysis or training a machine learning model.

The following are some common steps in data cleaning:

1. Handling missing data: Missing data can be imputed using various techniques such as mean, median, mode, or using machine learning algorithms to predict the missing values.
2. Removing duplicate records: Duplicates records can lead to bias in the analysis, and hence it's crucial to remove them from the dataset.
3. Outlier detection and removal: Outliers are data points that lie far away from the other data points and can skew the analysis or model training. Hence, it's essential to detect and remove outliers.
4. Handling inconsistent data: Inconsistent data can arise due to human error, data entry errors, or merge errors. These errors can be corrected manually or using machine learning algorithms.
5. Data transformation: Data transformation techniques such as normalization, scaling, or encoding can be applied to make the dataset more suitable for analysis or model training.

Data cleaning helps to improve the quality of data and leads to better analysis or machine learning models.

**Categorical Encoding**

Categorical encoding is the process of converting categorical data into numerical data, which can be used as input for machine learning algorithms. Categorical data is a type of data that includes categories or labels that are not numerical in nature. For example, colors, names, or types of products.

There are various techniques used for categorical encoding, including one-hot encoding, label encoding, binary encoding, and target encoding.

One-hot encoding is a technique where each category is converted into a binary vector, where only one element is 1 and the others are 0. Label encoding is a technique where each category is assigned a unique numerical value. Binary encoding is a technique where categories are converted into binary digits, and target encoding is a technique that involves encoding the categories based on their relationship with the target variable.

The choice of categorical encoding technique depends on the nature of the data and the machine learning algorithm used for analysis or prediction. Categorical encoding is an important step in data preprocessing, as it can significantly affect the performance of the machine learning model.

**PCA[Principal Component Analysis]**

PCA (Principal Component Analysis) is a dimensionality reduction technique that is used to reduce the number of variables in a dataset while retaining as much information as possible. It works by identifying the most important features or components in the data and transforming the data into a new coordinate system that represents these components.

PCA can prevent overfitting by reducing the complexity of the model. Overfitting occurs when a model is too complex and fits the training data too closely, which can lead to poor generalization to new or unseen data. PCA reduces the number of features used in the model, which can help to reduce the complexity of the model and prevent overfitting.

PCA works by identifying the principal components in the data, which are the directions in which the data varies the most. These principal components are orthogonal to each other and represent the most important information in the data. By projecting the data onto these principal components, PCA can reduce the dimensionality of the data while retaining most of the information.