Market Segmentation Analysis (Summary of step 4)- Ayush Dudhe

A First Glimpse at the Data

In order to determine which method is best for market segmentation, exploratory data analysis, or EDA, cleans and preprocesses the data after it has been collected. In a technical sense, EDA facilitates the determination of variable measurement levels, the study of univariate distributions, and the evaluation of interdependencies. Additionally, the data is prepared at this stage for entry into segmentation algorithms. The EDA-obtained insights guide the suitability of various techniques for deriving significant market segments.

Customers' preferences and opinions of several food aspects, like flavour, convenience, healthiness, and pricing, were gathered, for instance, in order to compile data for the McDonald's dataset. The dataset has demographic data like age, gender, and frequency of visits, as well as columns with categorical answers like "Yes" or "No" for qualities like "yummy," "convenient," and "healthy." This methodical technique makes it possible to thoroughly analyse customer behaviour and preferences, which yields insightful information that can be used to enhance operational strategy and customer happiness.

Data cleaning

Finding and fixing mistakes, inconsistencies, and inaccuracies in datasets is known as data cleaning. It entails dealing with outliers, formatting errors, addressing missing numbers, eliminating duplicates, and standardising data. Before analysis, this phase is essential for guaranteeing data reliability and quality. The objective is to produce a clear, consistent dataset that faithfully captures the data so that further studies can produce more accurate insights and decisions.
To guarantee the accuracy and usability of the data, the McDonald's dataset underwent a number of important steps in data cleansing. To make analysis easier, categorical replies like "Yes" and "No" were initially converted to numerical values (1 and 0, respectively). The identification and correction of missing values and differences included handling any null entries in the demographic data and transforming textual representations of numerical data into float format. As part of this process, the formats for the age and visit frequency entries were likewise standardised throughout the dataset. These actions gave the dataset a clear, organised format that was ready for further analysis and visualisation.

Descriptive analysis

When it comes to market segmentation, descriptive analysis entails analysing and summarising the salient features of the various client groups that are discovered via segmentation strategies. Within each segment, this process usually involves computing statistical measures for different demographic, behavioural, and psychographic characteristics, such as means, medians, and standard deviations. Charts, graphs, and tables are examples of visualisation tools that analysts use to show data in a style that is easier to understand. Aspects including age distribution, income brackets, spending patterns, brand preferences, and lifestyle characteristics unique to each group are frequently included in the

analysis. Marketers can learn more about the distinct characteristics and requirements of each consumer group by comparing these descriptive statistics across segments. This allows them to better target their advertising, services, and product offerings. Businesses can make better decisions regarding product creation, price, advertising, and distribution channels by having a thorough understanding of segment characteristics. This, in turn, improves consumer satisfaction and targeting. Continuous variables may be modified or binned to better fit the segmentation strategy, while categorical variables may be encoded or recoded for consistency. Cleaning procedures for text data frequently involve eliminating special characters, adjusting case, and fixing typos. Lastly, analysts make sure all the factors make sense and sometimes even develop derived variables to offer more segmentation-related information. Producing precise, significant segments that accurately reflect the market structure requires a rigorous cleaning procedure.

Preprocessing

In market segmentation analysis, data cleaning includes addressing numerous data quality issues to prepare the information for accurate analysis. This procedure entails addressing data discrepancies, finding and addressing outliers, standardising variables to ensure comparability, deleting duplicate records, encoding categorical variables, and handling missing values by imputation or deletion. Text data may be cleaned to fix formatting issues and standardise information. The objective is to produce a clear, standardised dataset that fairly depicts the market and enables trustworthy segmentation findings and insights. Ensuring the legitimacy and efficacy of the market segmentation analysis that follows depends on this phase.

Numerical and categorical variables in data preprocessing need to be handled differently. Discrete categories or groups are represented by categorical variables, which frequently require encoding—such as label encoding or one-hot encoding—to transform them into a format that can be used for analysis. Scaling (normalisation, standardisation, etc.) may be necessary for numerical variables, which are quantitative measurements, in order to guarantee that each feature contributes equally to the analysis. If necessary, numerical variables may also be binned into categories. Managing missing values, which may entail imputation or removal, is essential for both kinds. For numerical variables in particular, outlier detection and handling are crucial. With the goal of maximising the performance of ensuing analytical models, the preprocessing approach selected will rely on the particular analysis method and the characteristics of the data.

Principle component analysis

In market segmentation, Principal Component Analysis (PCA) is a dimensionality reduction technique that reduces the complexity of complicated datasets while maintaining the majority of the original information. It functions by locating the primary axes (also known as principal components) that exhibit the most variation in the data. These elements are ranked according to how much variance they can explain and have no correlation. PCA aids in the reduction of several factors (such as consumer traits or behaviours) in market

segmentation to a smaller set of composite variables that highlight the most significant patterns in the data. This facilitates the analysis and visualisation of the data as well as the identification of the critical elements that distinguish different client categories. Marketers can more effectively segment their market and create customised plans based on the most important client attributes by concentrating on these key elements.