**Market Segmentation Analysis - Summary Report**

***Keerthana J***

***Mahesh D***

***Ajay Kumar Lakkakula***

***Harshith Raj Krishna***

***Monali Jadhav***

Date: 26th April 2024

**What is Market Segmentation?**

Market segmentation is the process of dividing a broad target market into smaller, more defined segments based on certain characteristics or criteria. The purpose of segmentation is to identify groups of customers who share similar needs, preferences, behaviors, or characteristics, allowing businesses to tailor their marketing strategies and offerings to better meet the needs of each segment.

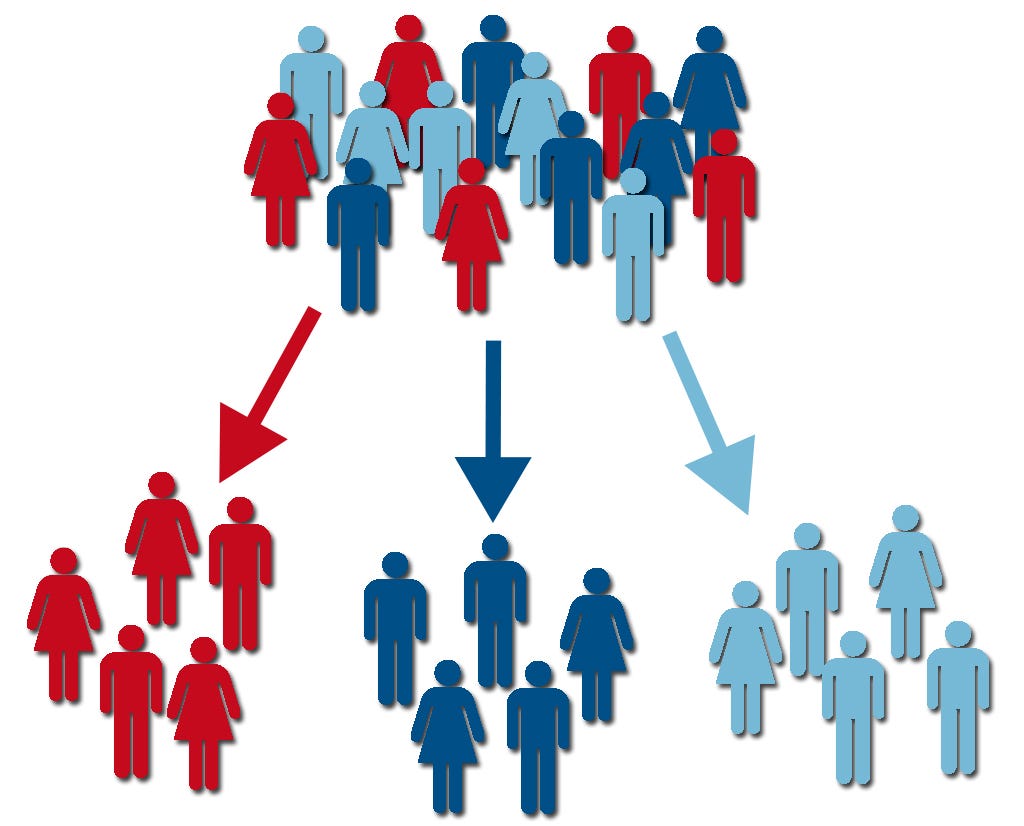


Figure 1: Understanding market segmentation.

<https://images.app.goo.gl/cBRNXEyDBwqHHemg6>

Key aspects of market segmentation include:

1. Demographic Segmentation: Dividing the market based on demographic factors such as age, gender, income, occupation, education, marital status, and family size.

2. Psychographic Segmentation: Categorizing consumers based on their lifestyles, attitudes, values, interests, personality traits, and behaviour patterns.

3. Behavioural Segmentation: Segmenting the market according to customer’s buying behaviour, usage patterns, brand loyalty, purchase frequency, benefits sought, and attitudes towards products or services.

4. Geographic Segmentation: Grouping consumers based on their geographic location, such as country, region, city, climate, population density, or urban-rural classification.



Figure 2: Types of Market Segmentation

<https://images.app.goo.gl/uPqqjtXvkd4rrLtx7>

Market segmentation enables businesses to:

* Identify and understand their target customers more effectively.
* Develop tailored marketing strategies, products, and services for each segment.
* Improve customer satisfaction and loyalty by meeting specific needs and preferences.
* Allocate resources more efficiently by focusing on high-potential segments.
* Gain a competitive advantage by offering differentiated value propositions.
* Enhance marketing effectiveness and ROI by delivering targeted messages to specific segments.
* Adapt to changing market dynamics and consumer trends more efficiently.

Overall, market segmentation is a strategic approach that helps businesses better understand their customers and create more relevant and impactful marketing initiatives, ultimately driving business growth and profitability.

**Step 1: Deciding (not) to Segment**

The step provides a detailed framework for evaluating market segments, stressing the crucial involvement of users throughout the segmentation process. It emphasizes that user input should be integral, spanning from initial briefings to final marketing mix development. Two distinct sets of criteria are outlined: knock-out criteria and attractiveness criteria. Knock-out criteria, including homogeneity, distinctiveness, size, organizational compatibility, identifiability, and reachability, are deemed essential for segment consideration. Conversely, attractiveness criteria encompass factors like profitability, accessibility, stability, and compatibility, assessing the desirability of remaining segments. Implementing a structured process for segment evaluation is recommended, leveraging tools such as segment evaluation plots to visualize segment attractiveness and organizational competitiveness. Collaboration among team members and stakeholders from various organizational units is crucial for defining and weighting criteria, with consensus-building facilitated through iterative discussions. Ultimately, the objective is to identify the most suitable target segments, aligning organizational goals with market realities through comprehensive evaluation and user engagement.

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

In Step 2 of market segmentation, the organization plays a crucial role by contributing to the determination of segment evaluation criteria. These criteria serve as guiding principles for identifying the most suitable target segments. The first set of criteria, known as knock-out criteria, are indispensable features that segments must possess to be considered viable options for targeting. These include homogeneity, where segment members should share similar characteristics, distinctiveness, indicating that the segment is different from others, and size, ensuring that it contains enough consumers to justify customized marketing efforts. Additionally, compatibility with the organization's strengths and capabilities, identifiability in the marketplace, and reachability for effective communication are essential criteria.

In contrast, attractiveness criteria provide a more nuanced evaluation of segments based on various factors that contribute to their appeal. These criteria encompass a wide range of considerations, including profitability, stability, accessibility, compatibility with the company's strengths and image, relationships with other segments, and alignment with corporate objectives. The selection of attractiveness criteria should reflect the specific needs and goals of the organization, with input from stakeholders across different organizational units to ensure comprehensive coverage and diverse perspectives.

While the segment evaluation plot, a commonly used structured approach, cannot be completed in this step, defining attractiveness criteria early in the process facilitates subsequent data collection efforts and streamlines the selection of target segments in later stages. This process involves assigning weights to each criterion to indicate its relative importance, with input from team members and approval from the advisory committee. By engaging stakeholders from various organizational units, the organization can benefit from a holistic perspective and ensure that the chosen criteria align with its overall objectives and capabilities.

**Step 3: Collecting Data**

The process of collecting data for market segmentation involves identifying segmentation variables and selecting suitable criteria for segmentation. Segmentation variables are the characteristics used to split the sample into market segments, while segmentation criteria determine the nature of the information used for segmentation.

In commonsense segmentation, a single characteristic such as gender is used as the segmentation variable, while in data-driven segmentation, multiple variables are utilized. Descriptor variables are additional characteristics used to describe segments in detail.

Segmentation criteria include geographic, socio-demographic, psychographic, and behavioural factors. Geographic segmentation relies on consumers' location of residence, making it easy to target communication messages but may not capture relevant characteristics for all products. Socio-demographic segmentation criteria include age, gender, income, and education, which can be useful for certain industries but may not explain specific product preferences. Psychographic segmentation focuses on psychological criteria like beliefs, interests, and preferences, offering deeper insights into consumer behaviour but requiring more complex analysis. Behavioural segmentation considers actual behaviour or reported behaviour, providing insights directly related to consumer actions.

Survey data is commonly used for segmentation studies, but it can be influenced by biases. Careful selection of variables and response options is crucial to ensure data quality and validity. Binary or metric response options are preferable for segmentation analysis, as they allow for straightforward distance measures.

* Response Styles: Survey data can be influenced by response biases, leading to response styles that affect the interpretation of segmentation results. For example, some respondents may consistently agree with all questions, creating a bias that skews segmentation analysis. It's crucial to identify and minimize response styles to ensure accurate segmentation results.
* Sample Size: The adequacy of sample size significantly impacts the effectiveness of segmentation analysis. Insufficient sample size makes it challenging to determine the correct number and nature of segments. Recommendations vary, but generally, a sample size of at least 60-70 times the number of segmentation variables is suggested for accurate results.
* Internal Data Sources: Organizations increasingly utilize internal data, such as scanner data or online purchase records, for segmentation analysis. These data offer insights into actual consumer behaviour but may be biased toward existing customers. It's essential to consider the representativeness of internal data for comprehensive segmentation.
* Experimental Studies: Experimental data, derived from field or laboratory experiments, provide valuable insights for segmentation analysis. For instance, responses to advertisements or preferences in choice experiments can serve as segmentation criteria. Conjoint studies offer information on how specific product attributes influence consumer choice, aiding segmentation efforts.

Overall, the quality of empirical data is essential for developing effective segmentation solutions, whether through commonsense or data-driven approaches. The choice of segmentation criteria should align with the organization's marketing goals and the nature of the product or service being analysed.

**Step 4: Exploring Data**

**(Ajay Kumar)**

The chapter provides a comprehensive guide to exploring and preparing data for market segmentation analysis after the data collection phase. This critical step ensures the data is clean, well-understood, and properly formatted before attempting to extract meaningful market segments. Several key processes are covered:

**Data Cleaning**: This crucial first step involves thoroughly examining the data for any errors, inconsistencies, or implausible values that may have been introduced during data collection or entry. All values are checked against permissible ranges and categories to identify and correct any issues. Consistent labelling of categorical variable levels is also ensured. Proper data cleaning lays a solid foundation for accurate analysis.

**Descriptive Analysis:** To gain familiarity with the data, extensive descriptive analysis is conducted using numerical summaries and visualizations. Histograms, boxplots, bar charts and other graphical methods are employed to understand the distribution, central tendency, dispersion, and overall characteristics of the variables, both numeric and categorical. This deep exploration reveals insights into the data structure and heterogeneity across respondents.

**Preprocessing of Variables:** Depending on the nature of the variables, different preprocessing steps may be applied. For categorical variables, this could involve merging granular categories for better interpretability or converting them to numeric format if treating them as ordinal scales is justified. For numeric variables, standardization may be performed to put them on a common scale and balance their influence in distance-based segmentation methods.

**Principal Components Analysis (PCA):** PCA is a dimensionality reduction technique that transforms the data into new uncorrelated variables called principal components, ordered by the amount of variability they explain. Though using only, a subset of top principal components as segmentation variables is inadvisable, PCA serves as a valuable exploratory tool. It helps identify redundant, highly correlated variables that may be capturing similar information. The principal component loadings provide insights into the underlying data structure. The chapter walks through the application of these processes using a real-world travel motives dataset, illustrating how rigorous data exploration guides the selection of appropriate segmentation methods and validates the suitability of the segmentation variables based on the heterogeneity observed across respondents. Clear, reproducible code is provided. In summary, this chapter underscores the critical importance of thoroughly exploring and preparing data through cleaning, descriptive analysis, preprocessing, and dimension reduction techniques like PCA. These steps ensure a solid foundation for accurate and insightful market segmentation in subsequent phases.

**Step 5: Extracting Segments**

**(Monali Jadhav and Harshith Krishna)**

1. **Grouping Consumers**

Grouping consumers in the context of market segmentation involves dividing a larger market into smaller, more homogeneous groups based on certain characteristics. These characteristics could include demographics, psychographics, behaviors, or other relevant factors. The goal is to identify distinct groups of consumers with similar needs, preferences, and behaviors so that marketing strategies can be tailored to each segment. Once consumers are grouped into segments, businesses can develop marketing strategies that are more relevant and effective for each segment. This can lead to improved customer satisfaction, increased brand loyalty, and a better understanding of the unique needs and preferences of various consumer groups.

1. **Distance-Based Methods**

Distance-based methods are a class of techniques used in data analysis and clustering to measure the similarity or dissimilarity between objects. These methods play a crucial role in clustering algorithms, where the goal is to group similar objects together. These distance measures are fundamental in various applications, including clustering, classification, and similarity-based searches. The choice of distance measure depends on the nature of the data and the specific requirements of the analysis. Different distance measures may yield different results, and selecting an appropriate measure is crucial for the success of the analysis. Distance Measures Each row represents an observation (in this case a tourist), and every column represents a variable (in this case a vacation activity). Mathematically, this can be represented as an n × p matrix where n stands for the number of observations (rows) and p for the number of variables (columns) The vector corresponding to the i-th row of matrix X is denoted as xi = (xi1, xi2,...,xip) in the following, such that X = {x1, x2,... xp} is the set of all observations. In the example above, Anna’s vacation activity profile is vector x1 = (100, 0, 0) and Tom’s vacation activity profile is vector x7 = (50, 20, 30).

2.1 Hierarchical Methods

Hierarchical clustering methods are the most intuitive way of grouping data because they mimic how a human would approach the task of dividing a set of n observations (consumers) into k groups (segments). Each consumer represents their own cluster. Market segmentation analysis occurs between those two extremes. Divisive hierarchical clustering methods start with the complete data set X and splits it into two market segments in a first step. Then, each of the segments is again split into two segments. This process continues until each consumer has their own market segment.

Agglomerative hierarchical clustering approaches the task from the other end.

The starting point is each consumer representing their own market segment (n singleton clusters). Step-by-step, the two market segments closest to one another are merged until the complete data set forms one large market segment. Both approaches result in a sequence of nested partitions. A partition is a grouping of observations such that each observation is exactly contained in one group. The sequence of partitions ranges from partitions containing only one group (segment) to n groups (segments). They are nested because the partition with k + 1 groups (segments) is obtained from the partition with k groups by splitting one of the groups.

2.2 Partitioning Method

Hierarchical clustering methods are particularly well suited for the analysis of small data sets with up to a few hundred observations. For data sets containing more than 1000 observations (consumers), clustering methods creating a single partition are more suitable than a nested sequence of partitions. This means that – instead of computing all distances between all pairs of observations in the data set at the beginning of a hierarchical partitioning cluster analysis using a standard implementation. A partitioning clustering algorithm aiming to extract five market segments, in contrast, would only have to calculate between 5 and 5000 distances at each step of the iterative or stepwise process (the exact number depends on the algorithm used). In addition, if only a few segments are extracted, it is better to optimise specifically for that goal, rather than building the complete dendrogram and then heuristically cutting it into segments.

2.3 Hybrid Approaches

Several approaches combine hierarchical and partitioning algorithms in an attempt to compensate the weaknesses of one method with the strengths of the other. The strengths of hierarchical cluster algorithms are that the number of market segments to be extracted. The biggest disadvantage of hierarchical clustering algorithms is that standard implementations require substantial memory capacity, thus restricting the possible sample size of the data for applying these methods. The basic idea behind hybrid segmentation approaches is to first run a partitioning algorithm because it can handle data sets of any size.

But the partitioning algorithm used initially does not generate the number of segments sought. Rather, a much larger number of segments is extracted. Then, the original data is discarded and only the centres of the resulting segments (centroids, representatives of each market segment) and segment sizes are retained, and used as input for the hierarchical cluster analysis. At this point, the data set is small enough for hierarchical algorithms, and the dendrogram can inform the decision how many segments to extract.

1. **Model-Based Methods**

Model-based methods for extracting segments refer to clustering techniques that involve fitting probabilistic models to the data. Unlike distance-based methods, which focus on measuring similarity or dissimilarity between data points, model based methods assume that the data is generated from a certain underlying probabilistic model. These methods aim to find the parameters of the model that best explain the observed data and, in the context of clustering, identify distinct segments or clusters within the data. One commonly used model-based clustering method is the Gaussian Mixture Model (GMM). GMM assumes that the data is generated from a mixture of several Gaussian distributions. Each Gaussian distribution represents a cluster, and the model estimates the parameters of these distributions, including mean, covariance, and mixing coefficients. Model-based clustering methods provide a flexible framework for segment extraction, allowing for more complex data distributions compared to some distance-based methods. It's important to understand the assumptions of the chosen model and validate the results based on the characteristics of the data and the goals of the analysis.

1. **Algorithms with Integrated Variable Selection**

These algorithms assume that each of the segmentation variables makes a contribution to determining the segmentation solution. Sometimes, segmentation variables were not carefully selected and contained redundant or noisy variables. Preprocessing methods can identify them. Variable selection for binary data is more challenging because single variables are not informative for clustering, making it impossible to pre-screen or pre-filter variables one by one. When the segmentation variables are binary, and redundant or noisy variables cannot be identified and removed during data pre-processing, suitable segmentation variables need to be identified during segment extraction. Several algorithms extract segments while–simultaneously– selecting suitable segmentation variables.

1. **Data Structure Analysis**

Extracting market segments is inherently exploratory, irrespective of the extraction algorithm used. Validation in the traditional sense, where a clear optimality criterion is targeted, is therefore not possible. Ideally, validation would mean calculating different segmentation solutions, choosing different segments, targeting them, and then comparing which leads to the most profit, or most success in mission achievement. As a consequence, the term validation in the context of market segmentation is typically used in the sense of assessing reliability or stability of solutions across repeated calculations after slightly modifying the, or the This approach is fundamentally different from validation using an external validation criterion. Throughout this book, we refer to this approach as stability-based data structure analysis. Data structure analysis provides valuable insights into the properties of the data. These insights guide subsequent methodological decisions. Most importantly, stability-based data structure analysis provides an indication of whether natural, distinct, and well-separated market segments exist in the data or not. If they do, they can be revealed easily. If they do not, users and data analysts need to explore a large number of alternative solutions to identify the most useful segment(s) for the organisation. If there is structure in the data, be it cluster structure or structure of a different kind, data structure analysis can also help to choose a suitable number of segments to extract.

**Step 6: Profiling Segments**

**(Ajay Kumar)**

This chapter delves into the crucial task of profiling and interpreting the market segments obtained through data-driven segmentation techniques. Profiling involves uncovering the distinct characteristics that define and differentiate each segment based on the variables used for segmentation. This process is vital for transforming the segmentation results into actionable insights that can guide marketing strategies.

The chapter highlights the limitations of traditional approaches used for presenting segment profiles. One approach involves oversimplified summaries that fail to capture the nuances of segment characteristics. The other approach relies on complex tables displaying statistics (percentages, means) for each segmentation variable across all segments. However, such tabular formats are cumbersome to interpret, requiring extensive comparisons between segments and against the overall sample.

To address these limitations, the chapter advocates the use of graphical data visualization techniques, which can effectively convey segment profiles in an intuitive and easy-to-comprehend manner. Specifically, it introduces two key visualizations: segment profile plots and segment separation plots. Segment profile plots provide a visual representation of how each segment differs from the overall sample across all segmentation variables. Each segment is represented by a panel, with distinctive "marker" variables that characterize the segment highlighted using color-coding. This visual format eliminates the need for tedious comparisons across segments and against the total sample. Segment separation plots aim to visualize the extent of overlap or separation between segments within the multi-dimensional data space defined by the segmentation variables. This is achieved through dimensionality reduction techniques like principal component analysis, which project the high-dimensional data onto two or three dimensions for plotting.

These plots illustrate the spread and shapes of segments, along with neighbourhood graphs indicating similarities between segments. The chapter provides examples demonstrating how these visualizations enable users to quickly grasp the defining features of each segment, unlike complex tables that demand significant cognitive effort. An eye-tracking study further reinforces the advantage of graphical formats, which require less mental effort from users to interpret segmentation solutions compared to tabular formats. Ultimately, the chapter underscores the value of effective data visualization techniques in facilitating a deep understanding of complex, multi-dimensional segmentation solutions. Well-designed graphical representations aid accurate interpretation of segment profiles, enabling managers to make informed strategic marketing decisions aligned with the identified market segments.

**Step 7: Describing Segments**

**(Keerthana J)**

7.1. Developing Complete Picture of Market Segmentation

Segmentation and its profiling is finding differences in different variables of the segment. The descriptor variables are used to profile different segments and to increase our understanding of the market more rather than just segregation of the market.

7.2. Visualizations for Description of Market Segmentation

Using visualizations such as charts and graphs to understand and draw insights from the empirical data so that the market as a whole can be understood. Trends and impact of different variables can be gauged in this step. Ordinal Descriptor Variables Use of descriptor variables are required to describe the market and effective segmentation. The use of segment number and segment plot will effectively project the different variables and the number of variables that are influencing the segment. Metric Descriptor Variables Use of metric variables with continuous values and trends of the variables is key to the segmentation and its success. The use of box plots, regression plots, graphs etc. are required to understand the data as a whole.

7.3 Testing for Segment Variables

The use of statistical methods and test are required for checking the success of the segmentation. Use of p and t test and sample distributions and probability theory are key tools and knowledge required to check it. The testing of model success using f1 score and R score will also be useful for successful segmentation.

7.4 Predicting Segments using Variables.

We need to develop different models using different approaches like regression classifier, decision trees, etc. The models would predict segments based on the different variables given as input and would be tested against the sample test data. The different models would be tested against each other on their relative success and the scores would determine the model that shall be used. Binary Logistic Regression We can make a model using binary classification variables in the data, the sample data will have to have binary descriptor variables and the model would be based on linear models and needs to be generalised to predict the segments. Multinomial Logistic Regression The use of logistic regression using different Metric variables and the combination of such variables is required to develop a model like this. The model requires generalisation of all the variables, and we need to be cautious not overfit it with variables and the generalisation. Tree Based Methods The use of tree on basis of nominal descriptor or metric descriptor would be a good start to generating a model, but the combination of the two different models would be ideal and should be used after normalisation and checking for overfitting. The development with the help of statistical method and random forests would ensure proper segmentation.

**Step 8: Selecting (the) Target Segment(s)**

**(Mahesh D)**

In the market segmentation process, selecting the target segment(s) emerges as a critical decision with far-reaching implications for the organization's future performance. This phase represents a pivotal moment akin to committing to a long-term relationship, where the organization must carefully weigh various market segments to identify the most suitable ones for its marketing efforts. Building on earlier groundwork, the focus shifts towards evaluating each segment against predetermined criteria, including factors like segment attractiveness and organizational competitiveness. Through this rigorous assessment, the segmentation team aims to pinpoint segments that not only align with the organization's strategic objectives but also offer significant growth potential and opportunities for sustainable success. Using decision matrices and visualization tools, such as the Boston or General Electric/McKinsey matrices, the team gains insights into the relative appeal of different segments and the organization's ability to thrive within them. Factors like market size, growth prospects, and the organization's capacity to meet segment needs are carefully considered in this process. Ultimately, the goal is to identify target segments that leverage the organization's strengths and provide avenues for meaningful engagement and growth. By selecting segments that offer the best fit with its capabilities and aspirations, the organization lays the foundation for long-term success in the marketplace.

**Step 9: Customising the Marketing Mix**

**(Mahesh D)**

The market segmentation process is intricately linked with other strategic marketing components, particularly positioning and competition. It operates within the broader framework of the segmentation-targeting-positioning (STP) approach, which emphasizes a sequential process starting with segmentation, followed by targeting, and culminating in positioning. However, while segmentation is often considered the initial step, the process is not strictly linear, and iterations between segmentation and targeting may be necessary before committing to long-term target segments. This integrated approach ensures that segmentation is not viewed in isolation but rather as part of a holistic strategy aimed at maximizing market potential and organizational success. Customizing the marketing mix to suit the selected target segment(s) is essential for optimizing the benefits of a market segmentation strategy. This involves thorough review and adaptation of each aspect of the marketing mix-Product, Price, Place, and Promotion-based on the characteristics and preferences of the target segment. For instance, product design or modification may be driven by segment needs, while pricing decisions should reflect the segment's willingness to pay. Similarly, distribution channels and promotional strategies should be tailored to effectively reach and engage the target segment. By aligning the marketing mix with the specific requirements of the chosen segments, organizations can enhance their competitive advantage and better meet customer needs, thereby driving long-term success in the marketplace.

**Case Study- Mc Donald’s**

A case study on market segmentation using McDonald's data in Python typically involves analysing customer demographics, purchasing behaviours, and preferences to identify distinct customer segments.

* **Data Collection**: The first step involves gathering relevant data from McDonald's, such as transaction records, customer demographics (age, gender, location), purchase history, and menu preferences. This data can be obtained from various sources, including internal databases or publicly available datasets.
* **Data Preprocessing**: Once the data is collected, it needs to be pre-processed to clean and prepare it for analysis. This may involve handling missing values, removing duplicates, and encoding categorical variables.
* **Exploratory Data Analysis (EDA)**: EDA is performed to gain insights into the characteristics and patterns present in the data. This includes visualizing distributions, correlations, and trends using techniques such as histograms, scatter plots, and heatmaps.
* **Market Segmentation**: Using machine learning or clustering algorithms such as K-means clustering or hierarchical clustering, the dataset is segmented into distinct customer groups based on similarities in their purchasing behaviours or demographics. These segments represent different market segments that McDonald's can target with tailored marketing strategies.
* **Segment Profiling**: Once the segments are identified, each segment is profiled to understand its unique characteristics, preferences, and behaviours. This involves analysing key metrics such as average order value, frequency of visits, popular menu items, and demographic profiles.
* **Marketing Strategy Recommendations**: Based on the segment profiles, recommendations are made for targeted marketing strategies to engage each segment more effectively. This may include personalized promotions, menu optimizations, or location-based marketing campaigns.
* **Evaluation and Iteration**: The effectiveness of the segmentation and marketing strategies is evaluated using metrics such as sales growth, customer satisfaction scores, and ROI. Feedback and insights from the evaluation phase are used to refine the segmentation approach and iterate on the marketing strategies.

**Implementation**

|  |  |  |
| --- | --- | --- |
| **Sno** | **Name** | **GitHub Link** |
| 1 | Keerthana J | [*https://github.com/keerthanaj2004/Feynn-Lab-Internship/tree/main/Task2*](https://github.com/keerthanaj2004/Feynn-Lab-Internship/tree/main/Task2) |
| 2 | Mahesh D | [*https://github.com/Mahesh050903/FeynnLab\_R-Py*](https://github.com/Mahesh050903/FeynnLab_R-Py) |
| 3 | Harshith Raj Krishna | [*https://github.com/Harshii69/feynn-lab-internship*](https://github.com/Harshii69/feynn-lab-internship) |
| 4 | Ajay Kumar Lakkakula | [*https://github.com/ajaykumarlakkakula/Python-code-*](https://github.com/ajaykumarlakkakula/Python-code-) |
| 5 | Monali Jadhav | [*https://github.com/Monalij28/FeynnLab\_McDonalds\_Case\_Study*](https://github.com/Monalij28/FeynnLab_McDonalds_Case_Study) |