**Market Segmentation – A Case Study**

*By*

Karan Singh

**Date: 8th July, 2023**



# Part I Introduction

## 1 Market Segmentation

* Strategic and Tactical Marketing
* Definitions of Market Segmentation
* The Benefits of Market Segmentation

## 2 Market Segmentation Analysis

* The Layers of Market Segmentation Analysis
* Approaches to Market Segmentation Analysis
* Based on Organizational Constraints
* Based on the Choice of (the) Segmentation Variable(s)
* Data Structure and Data-Driven Market Segmentation Approaches
* Market Segmentation Analysis Step-by-Step

# Part II Ten Steps of Market Segmentation Analysis

## Step 1: Deciding (not) to Segment

* Potentially required changes
* Implementation Barriers
* Second group and other barriers related to the segmentation

## Step 2: Specifying the Ideal Target Segment

* Segment Evaluation Criteria
* Knock-out criteria.
* Attractiveness Criteria.
* Implementing a Structured Process.

## Step 3: Collecting Data

* Segmentation Variables.
* Data from Survey Studies.
* Data from Internal Sources.

## Step 4: Exploring Data

* A First Glimpse at the Data
* Data Cleaning
* Descriptive Analysis
* Pre-Processing.

**Step 5: Extracting Segments**

## Step 6: Profiling Segments

* Identifying Key Characteristics of Market Segments
* Traditional Approaches to Profiling Market Segments
* Segment Profiling with Visualizations
* Identifying Defining Characteristics of Market Segments

**Step 7: Describing Segments**

## Step 8: Selecting Target Segment(s)

* The Targeting Decision
* Market Segment Evaluation

## Step 9: Customizing the Marketing Mix

* Implications of Marketing Mix Decisions.
* Profile and Detailed Description of the Target Segment • Product-Related Aspects for Target Segment
* Price-Related Aspects for Target Segment.
* Promotion-Related Aspects for Target Segment
* Place-Related Aspects for Target Segment
* Reviewing the Marketing Mix in its Entirety

**GitHub Links:**

# Part I. Introduction

**Market Segmentation**

## 1.1 Strategic and Tactical Marketing

Market segmentation involves both tactical and strategic marketing. To start, let's define market segmentation. Market segmentation is the practice of separating a large market into more manageable, smaller parts based on shared traits, requirements, or consumer behavior. Businesses can better understand their customers, target their marketing campaigns, and increase customer happiness and profitability by segmenting the market.

Now, let's delve into the concepts of strategic and tactical marketing in the context of market segmentation:

**Strategic Marketing:**

Strategic marketing plays an crucial role in shaping a company's long-term marketing activities. Within the context of market segmentation, strategic marketing focuses on several key aspects:

1. Target Market Selection: Strategic marketing involves carefully selecting the most suitable and attractive market segments that align with the company's overall objectives, available resources, and competitive advantages. This includes evaluating different segments based on factors such as their size, growth potential, profitability, and compatibility with the company's capabilities.
2. Positioning: Once the target segments are identified, strategic marketing determines how the company wants to position its products or services within each segment. Positioning entails creating a distinctive and favorable perception of the company's offerings compared to its competitors. It involves emphasizing the unique value proposition and key benefits that resonate with the specific needs and preferences of the target segment.
3. Resource Allocation: Strategic marketing also involves making effective resource allocation decisions to reach and serve the chosen market segments. This includes determining how to allocate the marketing budget, selecting the most suitable marketing channels, and distributing resources across different segments based on their strategic importance.

**Tactical Marketing:**

## Tactical marketing is the practical implementation of specific marketing strategies to effectively reach and engage with target market segments. It encompasses various key elements within the context of market segmentation:

## Message Customization: Tailoring marketing messages, content, and communication channels to align with the unique characteristics and needs of each market segment. This customization ensures that the marketing efforts are relevant and compelling to the specific segment, maximizing their impact.

## Channel Selection: Choosing the appropriate marketing channels to effectively reach the target segments. This involves considering the segment's media consumption habits, preferred communication channels, and overall marketing objectives to select the most suitable channels, whether traditional or digital, or a combination of both.

## Campaign Execution: Implementing marketing campaigns that are designed to engage with the target segments, generate leads, drive sales, and build brand awareness and loyalty. This includes activities such as advertising, promotions, content creation, events, and public relations, all aimed at effectively connecting with the identified segments.

## Performance Measurement: Monitoring and evaluating the performance of marketing initiatives to assess their effectiveness and make data-driven adjustments. (KPIs) such as customer acquisition, conversion rates, (ROI) are used to optimize future marketing efforts.

**1.2 Definitions of Market Segmentation**

## The process of splitting a large market into discrete groups or segments that have comparable traits, requirements, or behaviors is known as market segmentation. With the use of this technique, organizations may better understand their target market and target their marketing initiatives accordingly. Demographic, psychographic, geographic, and behavioral characteristics are used in segmentation. Businesses can more effectively manage resources, develop individualized marketing messages, and set themselves apart from rivals by focusing on certain segments. With this strategy, marketing initiatives are more successful, customers are more engaged, and overall business performance is enhanced.

## 1.3 The Benefits of Market Segmentation

* Better Understanding of Customers: Market segmentation provides businesses with a deeper understanding of their customers. By dividing the market into smaller segments, companies can identify and analyze specific customer groups with shared characteristics, needs, or behaviors.
* Improved Customer Satisfaction: When businesses understand their customers' needs and preferences through market segmentation, they can provide more relevant and personalized products, services, and experiences. By meeting the specific requirements of each segment, companies can enhance customer satisfaction levels.
* Increased Sales and Profitability: Targeted marketing through market segmentation can lead to increased sales and profitability. By focusing on the most receptive segments, businesses can generate higher conversion rates and customer acquisition.

# Part II Ten Steps of Market Segmentation Analysis

**Step 1: Deciding (Not) to segment**

It is crucial to comprehend the ramifications of pursuing a market segmentation strategy before devoting time and resources to a market segmentation analysis. The main takeaway is that the organization must make a long-term commitment to the segmentation strategy.

Potentially required changes are:

1. Development of new products.
2. Modification of existing products.
3. Changes in pricing and distribution channels.

The decision to study the possibility of a market segmentation strategy must be made at the highest executive level and must be consistently conveyed to and reinforced across all levels due to the significant consequences of such a long-term organizational commitment.

**Implementation Barriers:**

First group of barriers related to the market segmentation

* Lack of leadership,
* pro-active championing,
* commitment, and involvement in the market segmentation process by senior leadership undermines the success of market segmentation.

**Second group and other barriers related to the market segmentation:**

* Lack of training.
* Lack of a formal marketing function or atleast a qualified marketing expert in the organization.
* Lack of a qualified data manager and analyst in the organization.
* Lack of financial resources.
* Lack of planning or bad planning.
* Lack of structured processes.
* Lack of time to conduct the market segmentation analysis without time pressure.

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

In the second step of market segmentation analysis, the focus is on defining the target segment that best aligns with the organization's goals. This involves identifying and evaluating specific criteria to determine the suitability of each segment. The process requires input from the organization, collaboration among the segmentation team, and the advisory committee.

**Segment Evaluation Criteria:**

The evaluation criteria consist of two sets: knock-out criteria and attractiveness criteria.

**Knock-out criteria:**

Knock-out criteria are essential and non-negotiable factors that must be met for a segment to be considered as a target. These criteria include homogeneity, distinctness, size, match, identifiability, and reachability. If a segment fails to meet these criteria, it will be automatically excluded in the later stages of the analysis.

**Attractiveness Criteria:**

In contrast, attractiveness criteria are used to assess the relative appeal of the remaining market segments. These criteria are not binary but are evaluated on a scale. The literature provides a range of suggested attractiveness criteria, such as substantiality, measurability, accessibility, compatibility, profitability, and more. The segmentation team needs to select a subset of these criteria based on their relevance to the organization's specific situation. The importance or weight of each attractiveness criterion is determined through discussions and negotiations among team members.

**Step 3: Collecting Data**

Data collection, which entails acquiring essential information about potential clients and the market, is a key component of the market segmentation process. This stage is essential because it serves as the foundation for identifying separate client segments and recognising their particular traits. Numerous techniques and sources are frequently used to provide complete and accurate data.

Primary research, which involves direct interaction with the target market, is a popular method of data collection. Surveys, interviews, focus groups, or observations can all be used to this end. Businesses can gather precise and specialized data about customers' preferences, wants, behaviors, and demographics by connecting with them directly. Primary research offers first-hand knowledge that can be tailored to meet certain research goals.

On the other hand, secondary research is also an important method of data collection. It involves gathering existing data from diverse sources such as industry reports, government publications, market studies, and academic research. Secondary research offers a broader perspective on the market, industry trends, and customer segments. It helps to validate and complement the findings from primary research while providing a benchmark for comparison.

**Step 4: Exploring Data:**

**4.1. A First Glimpse at the Data**

After data collection, exploratory data analysis cleans and – if necessary – pre-processes the data. This exploration stage also offers guidance on the most suitable algorithm for extracting meaningful market segments. At a more technical level, data exploration helps to

⦁ identify the measurement levels of the variables

⦁ investigate the univariate distributions of each of the variables

⦁ assess dependency structures between variables.

Results from the data exploration stage provide insights into the suitability of different segmentation methods for extracting market segments.

The Data Exploration is key technique to understand the data in a better way.

The Data can be visualized by using the pie charts, bar charts, scatter plots, line graphs etc.

**4.2. Data Cleaning**

The first step before commencing data analysis is to clean the data. This includes checking if all values have been recorded correctly, and if consistent labels for the levels of categorical variables have been used.

The Data Cleaning may include the several steps:

⦁ Identify Data Quality Issues: The first step in data cleaning is to identify potential data quality issues. This includes examining the dataset for missing values, duplicate records, incorrect or inconsistent values, outliers, formatting errors, and other anomalies that may affect the integrity and reliability of the data.

⦁ Handle Missing Data: Missing data refers to the absence of values for certain variables or records in the dataset. Depending on the nature and extent of missing data, various techniques can be applied, such as imputation (estimating missing values based on existing data), deletion of records with missing values, or creating separate categories for missing data.

⦁ Remove Duplicates: Duplicate records occur when multiple entries with identical or highly similar data exist in the dataset. These duplicates can skew analysis results and lead to incorrect conclusions. Identifying and removing duplicates ensures that each record is unique and avoids bias in subsequent data analysis.

⦁ Standardize and Correct Inconsistent Values: Inconsistencies in data occur when different representations or formats are used for the same variable. For example, variations in date formats, spelling errors, or different units of measurement. Standardizing and correcting these inconsistencies involves transforming the data to a consistent format, resolving misspellings or abbreviations, and converting units of measurement to a common standard.

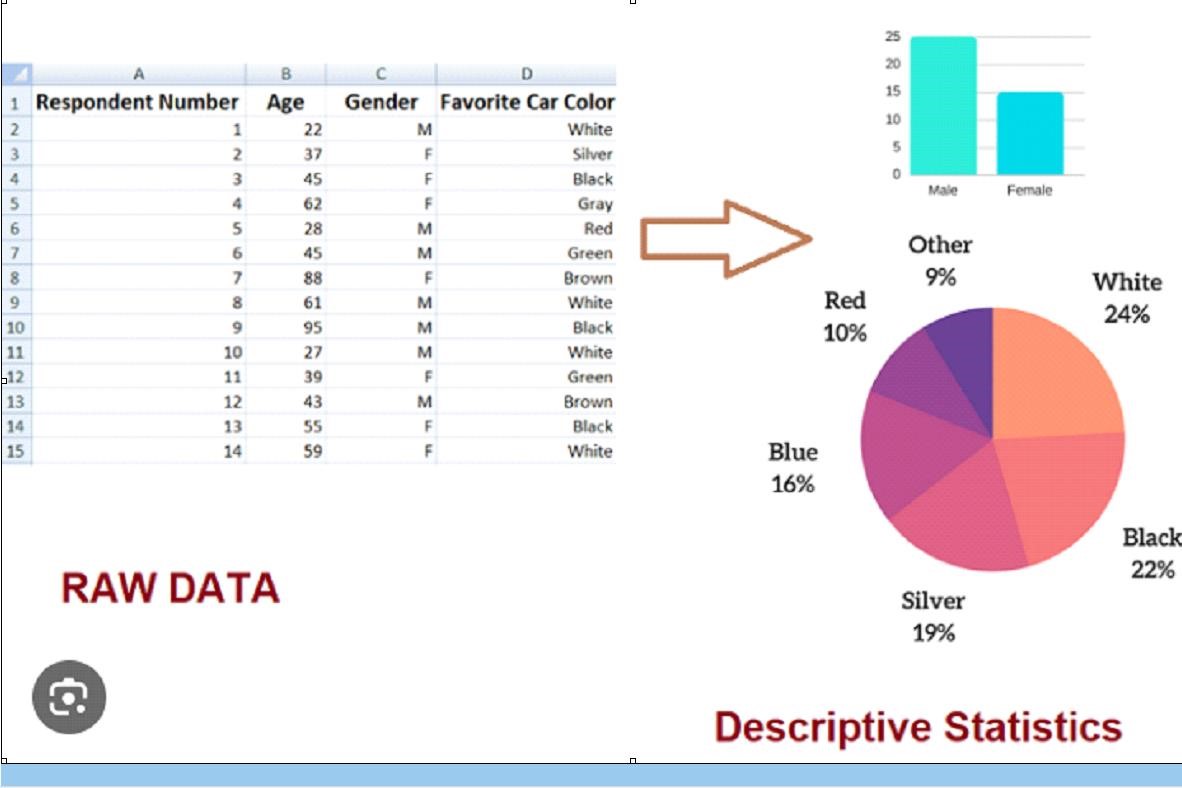
⦁ Handle Outliers: Outliers are data points that deviate significantly from most of the data. These extreme values can affect statistical analyses or modelling outcomes. Determining the source and nature of outliers is important to decide whether to remove them or treat them separately based on the context of the data and the analysis objectives.

**4.3. Descriptive Analysis**

Descriptive analysis in machine learning refers to the exploration and summarization of data to gain insights and understand the characteristics and patterns within the dataset. It involves using statistical and visualization techniques to describe the data's key features, distributions, relationships, and summary statistics. Descriptive analysis plays a crucial role in the initial stages of machine learning projects as it helps researchers and practitioners develop a deeper understanding of the data they are working with.

Some of the basic methods involving the Descriptive Analysis are:

* Summary Statistics: Summary statistics provide a concise overview of the dataset, including measures such as mean, median, mode, standard deviation, range, and quartiles. These statistics help understand the central tendency, variability, and distribution of the data.
* Data Distribution: Descriptive analysis allows examining the distribution of variables in the dataset. This involves analyzing the shape of distributions (e.g., normal, skewed), identifying any skewness or kurtosis, and checking for multimodality or presence of multiple peaks. Understanding the data distribution helps in selecting appropriate modelling techniques and identifying potential issues that may affect model performance.
* Correlation Analysis: Correlation analysis measures the strength and direction of relationships between variables. It helps identify whether variables are positively, negatively, or not correlated. Techniques such as correlation matrices or scatter plots can be used to visualize and quantify the relationships. Correlation analysis assists in understanding the dependencies and interactions between variables, which can guide feature selection or inform the modelling process.
* Missing Data Analysis: Descriptive analysis allows understanding the extent and patterns of missing data within the dataset. It involves analyzing missing data mechanisms (e.g., missing completely at random, missing at random, or missing not at random) and identifying potential patterns or biases associated with missing values. This analysis helps determine appropriate strategies for handling missing data, such as imputation or exclusion.



**4.4. Pre-Processing**

**4.4.1 Categorical Variables**

Preprocessing categorical variables is an important step in preparing data for machine learning algorithms. Categorical variables represent qualitative or non-numeric data, such as categories, labels, or groupings. Preprocessing these variables involves transforming them into a numerical format that can be easily understood and processed by machine learning algorithms.

Here are some common techniques for preprocessing categorical variables:

* **Label Encoding:** Label encoding assigns a unique numeric label to each category in a variable. Each category is mapped to a corresponding integer value. This technique is suitable for ordinal variables where the categories have a specific order or rank. However, label encoding may introduce unintended ordinality where none exists in the data.
* **One-Hot Encoding:** One-hot encoding, also known as dummy encoding, creates binary variables for each category in a categorical variable. Each category is transformed into a separate binary column, where a value of 1 indicates the presence of the category, and 0 indicates its absence. One-hot encoding is suitable for nominal variables without any inherent order or hierarchy.
* **Ordinal Encoding:** Ordinal encoding assigns numeric values to categories based on their relative order or rank. This technique is appropriate when the categories have an inherent order or if the variable is a combination of both nominal and ordinal values. Each category is assigned a unique numeric value based on its order or importance.
* **Target Encoding:** Target encoding, also known as mean encoding, replaces each category with the mean of the target variable for that category. It is particularly useful for classification problems, as it captures the relationship between the categorical variable and the target variable. However, target encoding may be prone to overfitting if not properly regularized.

**4.4.2. Numeric Variables**

Preprocessing numerical variables is an important step in preparing data for machine learning algorithms. While numerical variables are already in a numeric format, preprocessing helps to standardize or transform them to ensure they are suitable for the chosen machine learning algorithm.

Most popular techniques are discussed below:

⦁ **Standardization:** Standardization, also known as Z-score normalization, scales numerical variables to have a mean of 0 and a standard deviation of 1. It involves subtracting the mean of the variable from each data point and then dividing by the standard deviation. Standardization ensures that variables with different scales and units have a comparable range, which can be important for algorithms that are sensitive to scale.

⦁ **Min-Max Scaling:** Min-Max scaling transforms numerical variables to a specific range, typically between 0 and 1. It involves subtracting the minimum value from each data point and then dividing by the difference between the maximum and minimum values. Min-Max scaling preserves the relative relationships between data points and is useful when you want to constrain variables to a specific range.

⦁ **Feature Scaling:** Feature scaling ensures that all numerical variables are on a similar scale. It involves scaling variables to a specific range or unit magnitude. Feature scaling can be beneficial for algorithms that use distance-based calculations, such as clustering or k-nearest neighbors.

⦁ **Log Transformation:** Log transformation applies the natural logarithm to numerical variables. It is useful when the variable has a skewed distribution, and a log transformation can help make the data more symmetric. Log transformation can also help stabilize the variance of variables that exhibit heteroscedasticity.

**Step 5: Extracting Segments**

**5.1 GROUPING CONSUMERS**

Market segmentation analysis based on consumer data is exploratory and involves unstructured datasets. Consumer preferences are diverse and do not exhibit clear groups when plotted. Instead, preferences are scattered across the entire plot. Extracting market segments from such data heavily relies on the assumptions made about the segment structure. Thus, the outcome of a segmentation analysis is influenced by both the data itself and the chosen extraction algorithm. Segmentation methods play a significant role in shaping the resulting segmentation solution.

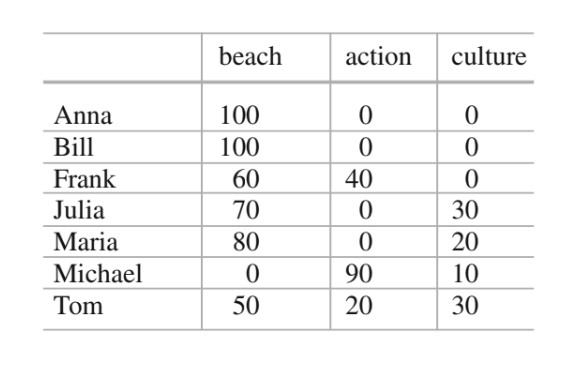
**FLOW OF GROUPING:**

1. Market segmentation analysis is exploratory.
2. Consumer datasets are often unstructured.
3. Consumer preferences do not exhibit clear groups when plotted.
4. Preferences are spread across the entire plot.
5. Extracting market segments depends on assumptions about segment structure.
6. Segmentation results are influenced by data and extraction algorithms.
7. Segmentation methods shape the segmentation solution.
8. Results depend on data, assumptions, and algorithms.
9. Consider limitations and biases when interpreting segmentation results.

**Data set and segment characteristics informing extraction algorithm selection**

|  |  |
| --- | --- |
| **Data set characteristics** | * Size (number of consumers, number of segmentation variables) * Scale level of segmentation variables (nominal, ordinal, metric, mixed) * Special structure, additional information |
| **Segment characteristics** | * Similarities of consumers in the same segment * Differences between consumers from different segments * Number and size of segments |

**5.2 DISTANCE BASED METHODS**



**PROBLEM STATEMENT**

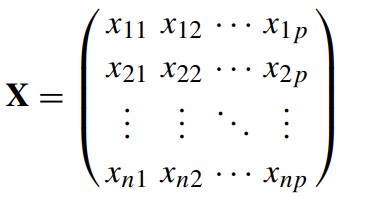
Grouping tourists based on similar activities and behavioral patterns.

**DATASET**

Artificial dataset was generated for seven tourists.

Market segmentation aims to group consumers based on their similar needs or behavior. In this case, the goal is to identify groups of tourists with similar vacation activity patterns. Anna and Bill have identical profiles and should be in the same segment since they exhibit the same characteristics. However, Michael stands out as he is the only one not interested in going to the beach, which sets him apart from the other tourists. To find these groups of similar tourists, a measure of similarity or dissimilarity is needed, typically represented mathematically as a distance measure. This distance measure helps quantify the differences between tourists and enables the identification of distinct segments based on their preferences and behaviors.

**DISTANCE MEASURES:**



In the given example, Anna's vacation activity profile is represented by the vector x1 = (100, 0, 0), while Tom's vacation activity profile is represented by the vector x7 = (50, 20, 30). These vectors represent observations in a matrix X.

To measure the distance between two vectors, various approaches exist and are commonly used in cluster analysis and market segmentation. A distance is a function, denoted as d(·, ·), with two arguments: the vectors x and y for which the distance is being calculated. The output is a nonnegative value representing the distance between the two vectors.

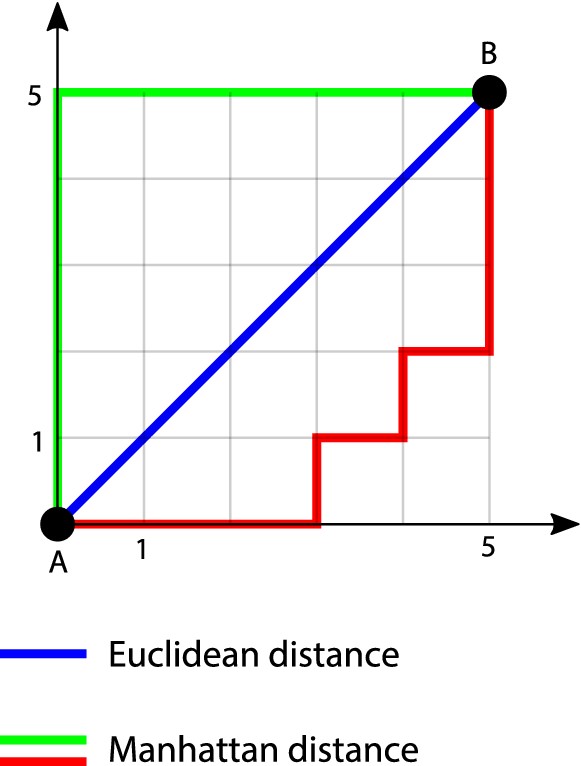
An analogy to understand distance is through geography. If we are interested in the distance between two cities, the vectors represent the locations of the cities, and the distance could be measured as the length of the air route between them in kilometers. However, there are other valid measures of natural distance between cities, such as the distance a car has to drive on roads to travel from one city to the other.

**FEW CRITERIA:**

* **d(x, y) = d(y, x)**
* **d(x, y) = 0** ⇔ **x = y.**
* **d(x, z) ≤ d(x, y) + d(y, z).**

**COMMON DISTANCE MEASURES USED:**

* Euclidean distance
* Manhattan distance
* Asymmetric binary distance



* The asymmetric binary distance is a distance measure used in market segmentation analysis.
* It only considers dimensions where at least one of the two vectors has a value of 1, ignoring dimensions with 0s in both vectors.
* It treats 0s and 1s differently, considering similarity only if the vectors share 1s but not if they share 0s.
* Dissimilarity between vectors increases if one vector has a 1 and the other does not.
* The asymmetric binary distance has implications for market segmentation analysis.
* Unusual activities, such as horseback riding and bungee jumping, may not contribute much to segment extraction if their overall proportions in the population are low.
* Commonalities in not engaging in certain activities are not informative for segment extraction.
* The proportion of common 1s over dimensions where at least one vector contains a 1 represents the asymmetric binary distance.
* In the tourist example, it is the number of common vacation activities divided by the number of vacation activities at least one of the two tourists engages in.
* A symmetric binary distance measure, treating 0s and 1s equally, can be obtained by using the Manhattan distance between the two vectors.
* In this case, the distance is equal to the number of vacation activities where the values differ.

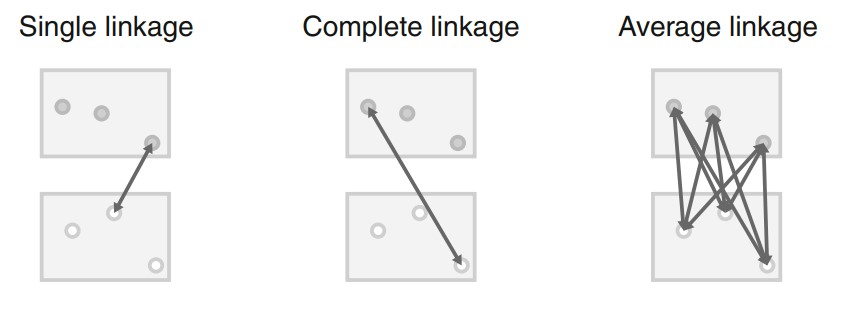
**HIERARCHICAL METHODS:**

Hierarchical clustering methods closely resemble how humans would naturally group data. They mimic the process of dividing a set of observations (consumers) into a desired number of groups (segments). When aiming for a single large market segment (k = 1), the only solution is to have one big segment that includes all consumers in the dataset. On the other hand, if the goal is to have as many market segments as there are consumers in the dataset (k = n), each consumer becomes their own individual segment, resulting in n segments. Market segmentation analysis typically falls between these two extremes, finding an optimal number of segments that strike a balance between homogeneity within segments and heterogeneity across segments.

* Divisive hierarchical clustering starts with the complete dataset and splits it into two segments, then continues to split each segment into further segments until each consumer has their own segment.
* Agglomerative hierarchical clustering starts with each consumer representing their own segment and gradually merges the two closest segments until the entire dataset forms one large segment.
* Both approaches result in a sequence of nested partitions, ranging from one group (segment) to n groups (segments).
* The partitions are nested because each partition with k+1 groups is obtained by splitting one group from the partition with k groups.
* Numerous algorithms have been proposed for both divisive and agglomerative clustering, with Lance and Williams' framework serving as a unifying approach for agglomerative clustering.
* Standard implementations of hierarchical clustering perform optimal steps in each iteration, resulting in a deterministic algorithm without random components.
* Each application of hierarchical clustering to the same dataset will yield the exact same sequence of nested partitions.

Assuming two sets X and Y of observations (consumers), the following linkage methods can be used:

* Single Linkage
* Complete Linkage
* Average Linkage



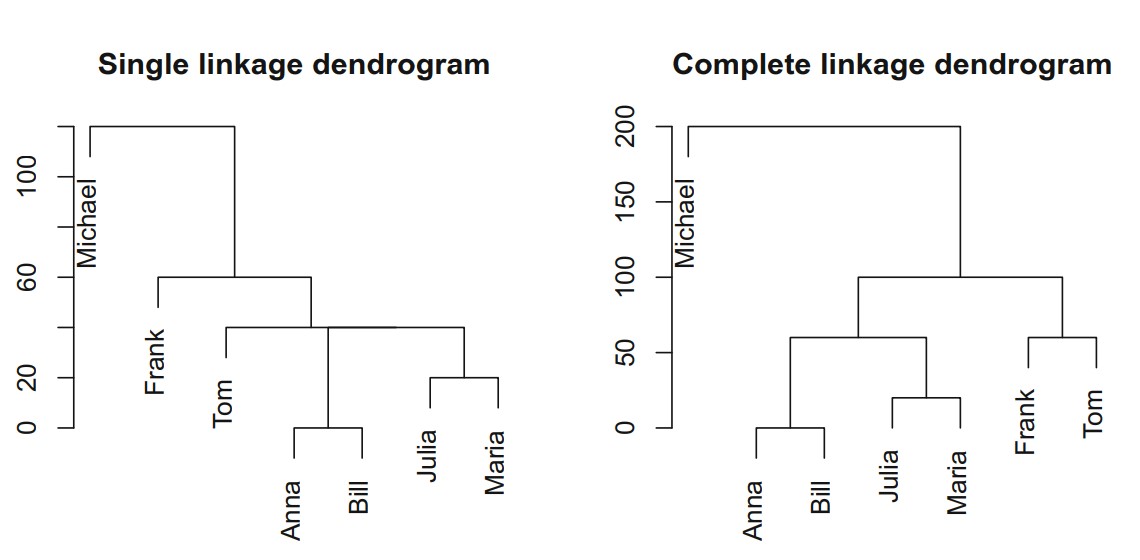
Both divisive and agglomerative clustering rely on a distance measure between groups of observations or segments. This distance measure is determined by specifying two factors: (1) a distance measure, denoted as d(x, y), between individual observations (consumers) x and y, and (2) a linkage method.

The distance measure, d(x, y), quantifies the dissimilarity or similarity between two individual observations. It provides a way to compare and assess the differences or similarities in their characteristics or attributes.

The linkage method, on the other hand, generalizes how distances between groups of observations are calculated based on the pairwise distances between individual observations within those groups. It determines how the distance between two groups is computed from the distances of their constituent observations.

By combining the distance measure and linkage method, both divisive and agglomerative clustering algorithms can determine the distances between groups of observations and iteratively merge or split segments based on these distances.

The result of hierarchical clustering is typically presented as a dendrogram. A dendrogram is a tree diagram. The root of the tree represents the one-cluster solution where one market segment contains all consumers. The leaves of the tree are the single observations (consumers), and branches in-between correspond to the hierarchy of market segments formed at each step of the procedure. The height of the branches corresponds to the distance between the clusters. Higher branches point to more distinct market segments. Dendrograms are often recommended as a guide to select the number of market segments.



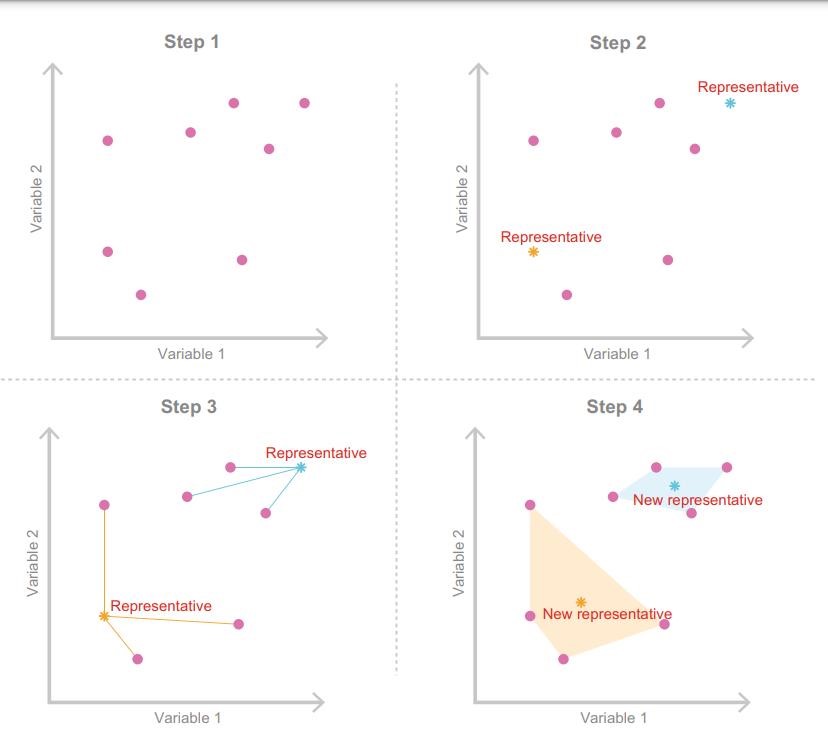
The observations or customers' order as tree leaves is not particular.

Every time there was a split into two branches, the left and right branches may be switched, giving rise to 2n alternative dendrograms for the exact same clustering, where n is the number of splits of consumers included in the data collection. Dendrograms generated by various software programs may therefore appear different even though they are identical in terms of the market segmentation they reflect. When numerous groups have exactly the same distance, how ties are broken—that is, which two groups are linked first—could be another possible source of variance among software packages.

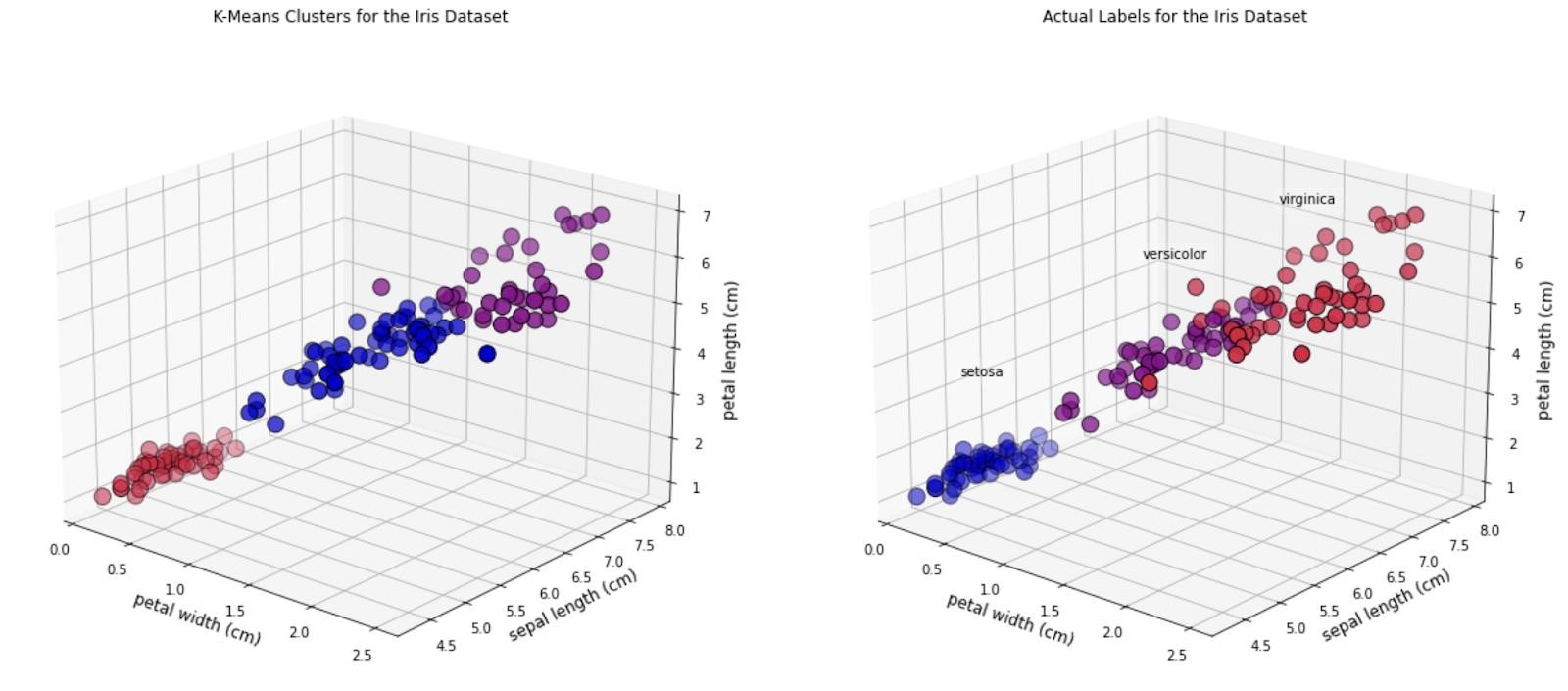
**PARTITIONING METHODS:**

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 optimize specifically for that goal, rather than building the complete dendrogram and then heuristically cutting it into segments.

* **K-Means Clustering**



* **K Centroid Clustering**

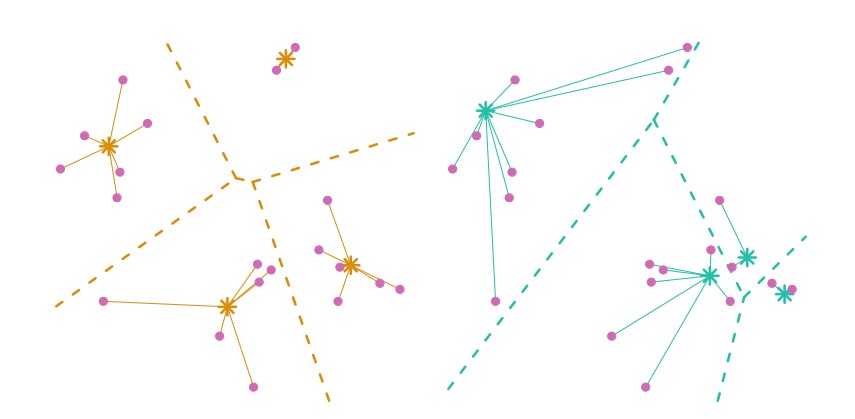


Several improvements have been proposed to enhance the k-means clustering algorithm. One straightforward improvement is to use "smart" initial values instead of randomly selecting k consumers from the dataset as starting points. Randomly selecting initial points can be suboptimal because some of them may end up being very close to each other, which does not represent the overall data space effectively. This increases the risk of the k-means algorithm getting stuck in a local optimum, a good solution but not the best possible one.

To avoid the issue of local optima, one approach is to initialize the algorithm with starting points evenly distributed across the entire data space. By using such evenly spread starting points, the algorithm can better capture the representation of the entire dataset and improve the chances of finding a globally optimal solution.

**HARD COMPETITIVE LEARNING**

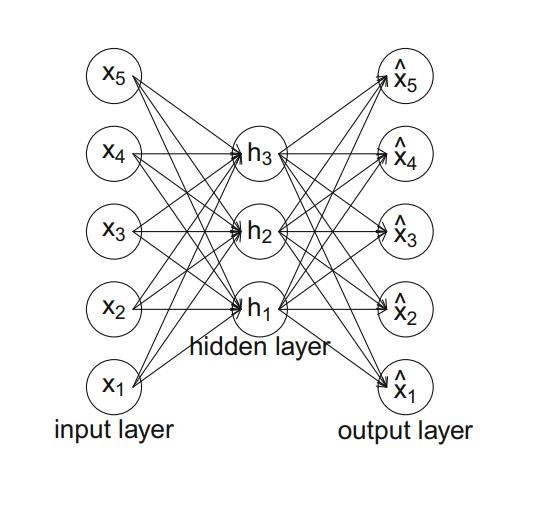
Hard competitive learning, also known as learning vector quantization (e.g. Ripley 1996), differs from the standard k-means algorithm in how segments are extracted. Although hard competitive learning also minimizes the sum of distances from each consumer contained in the data set to their closest representative (centroid), the process by which this is achieved is slightly different.



Starting point good and bad examples respectively.

**NEURAL NETWORKS**

* Auto-encoding neural networks are a unique approach to cluster analysis.
* They utilize a single hidden layer perceptron as the main method.
* The network consists of three layers: input, hidden, and output.
* The hidden layer is named so because it has no connections to the outside of the network.
* Each node in the hidden layer is a weighted linear combination of the inputs.
* The weights are represented by arrows connecting the input and hidden layers.
* Non-linear functions are used to determine the values of the hidden nodes.
* The outputs of the network are weighted combinations of the hidden nodes.
* The network is trained by adjusting the parameters to minimize the squared Euclidean distance between inputs and outputs.
* The training process aims to predict the inputs as accurately as possible.
* The number of hidden nodes is usually less than the number of inputs, forcing the network to learn how to represent the data using segment representatives.
* Once trained, the parameters connecting the hidden layer to the output layer are interpreted as segment representatives.
* The parameters connecting the input layer to the hidden layer indicate membership in different segments.
* Consumers with similar values in the hidden layer nodes are considered part of the same segment.



**HYBRID APPROACHES** :

* Two Step Clustering
* Bagged Clustering

**5.3 MODEL BASED METHODS:**

**FINITE MIXTURES OF DISTRIBUTIONS:**

* Normal Distributions
* Binary Distributions

**FINITE MIXTURES OF REGRESSIONS:**

* Linear Regression
* Lasso Regression

**EXTENSIONS AND VARIABLES:**

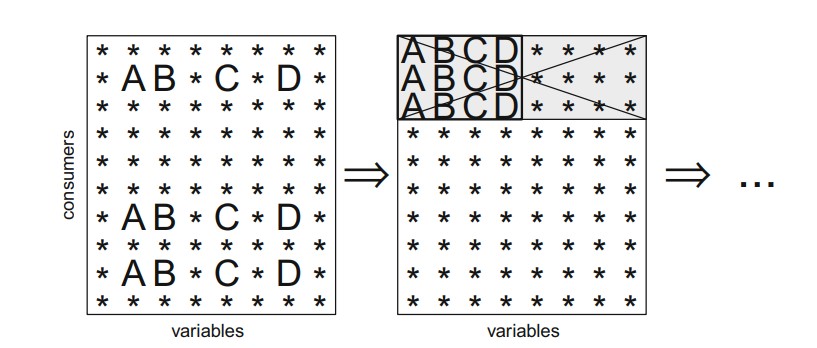
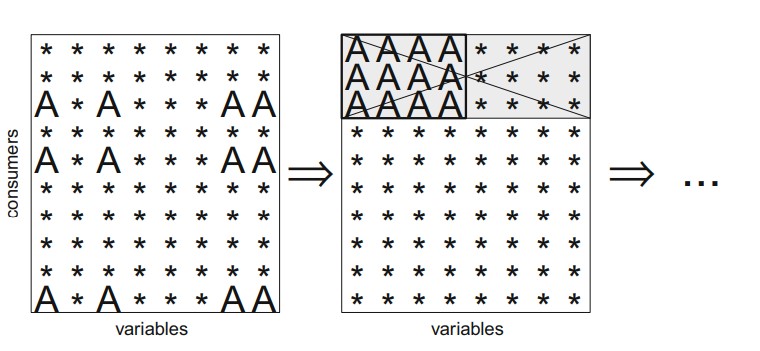
Finite mixture models are more complex than distance-based methods.

They offer flexibility by allowing the use of any statistical model to describe a market segment.

* Different data characteristics can be accommodated, such as metric data with mixtures of normal distributions and binary data with mixtures of binary distributions.
* Nominal variables can be handled with mixtures of multinomial distributions or multinomial logit models.
* Ordinal variables require special attention due to response styles, and mixture models can disentangle response style effects from content-specific responses.
* Mixture models combined with conjoint analysis can account for differences in preferences.
* There is a debate in the segmentation literature about whether to model differences between consumers using a continuous distribution or distinct market segments.
* Mixture models can reconcile these positions by recognizing the existence of distinct segments while allowing variation within each segment.
* Mixture of mixed-effects models or heterogeneity models can be used to model demand and capture both distinct segments and withinsegment variation.
* Mixture models can also be applied to time series data, clustering the time series and extracting groups of similar consumers.
* Dynamic latent change models, such as Markov chains, can track changes in brand choice and buying decisions over time.
* Mixture models can incorporate both segmentation and descriptor variables, with segmentation variables used for grouping and descriptor variables used to model differences in segment sizes.

**7.4 ALGORITHMS WITH INTEGRATED VARIABLE SELECTION**

* **BICLUSTERING ALGORITHMS**



* **VARIABLE SELECTION PROCEDURE FOR CLUSTERING BINARY DATA(VSBD):**

The algorithm works as follows:

* **Step 1:** Select only a subset of observations with size φ  (0, 1] times the size of the original data set. Brusco (2004) suggests to use φ = 1 if the original data set contains less than 500 observations, 0.2 ≤ φ ≤ 0.3 if the number 7.4 Algorithms with Integrated Variable Selection 149 of observations is between 500 and 2000 and φ = 0.1 if the number of observations is at least 2000.
* **Step 2:** For a given number of variables V , perform an exhaustive search for the set of V variables that leads to the smallest withincluster sum-of-squares criterion. The value for V needs to be selected small for the exhaustive search to be computationally feasible. Brusco (2004) suggests using V = 4, but smaller or larger values may be required depending on the number of clusters k, and the number of variables p. The higher the number of clusters, the larger V should be to capture the more complex clustering structure. The higher p, the smaller V needs to be to make the exhaustive search computationally feasible.
* **Step 3:** Among the remaining variables, determine the variable leading to the smallest increase in the within-cluster sum-ofsquares value if added to the set of segmentation variables.
* **Step 4:** Add this variable if the increase in within-cluster sumof-squares is smaller than the threshold. The threshold is δ times the number of observations in the subset divided by 4. δ needs to be in [0, 1]. Brusco (2004) suggests a default δ value of 0.5.
* **VARIABLE FACTOR ANALYSIS:**

In summary, the main points regarding factor-cluster analysis for market segmentation are as follows:

* Two-step procedure: Factor-cluster analysis is a data-driven market segmentation approach. In the first step, segmentation variables are factor analyzed, and the raw data is discarded. In the second step, market segments are extracted using factor scores obtained from the factor analysis.
* Conceptual legitimacy: Factor-cluster analysis can be conceptually legitimate when the original variables are replaced with factor scores that represent validated psychological test batteries or other variables designed to load onto factors. However, the factor scores should be determined simultaneously or separately from the data, not in a data-driven manner when groups are suspected.
* Sample size limitations: Factor-cluster analysis is often used when the number of segmentation variables is high compared to the sample size. Simulation studies suggest that the sample size should be at least 100 times the number of segmentation variables for reliable results. However, many market segmentation studies have a smaller sample size, making it challenging to meet this criterion.
* Information loss: Factor analysis leads to a substantial loss of information. The percentage of explained variance after factor analysis represents the amount of information retained. Using factor scores for segment extraction means sacrificing a significant portion of the information contained in the original segmentation variables.
* Data transformation: Factor analysis transforms the data, resulting in segments extracted from a modified version of the consumer data. This differs from extracting segments directly from the original segmentation variables, which may affect the representation and interpretation of market segments.
* Interpretation challenges: Factor-cluster results are more difficult to interpret compared to results based on the original segmentation variables. Factors contain partial information from multiple variables, making it challenging to translate segment profiles into practical recommendations for the marketing mix.

* Preference for raw data: Cluster analysis using raw data is often considered to produce more accurate or detailed segmentation results as it preserves a greater degree of the original data. Factor-cluster analysis is discouraged for market segmentation purposes, except when developing an instrument for the entire population assuming homogeneity among consumers.
* Performance compared to raw data clustering: Empirical evidence suggests that factor-cluster analysis does not consistently outperform clustering of raw data in terms of identifying the correct market segment structure. Even when the data is generated following a factor-analytic model, factor-cluster analysis may fail to capture the underlying structure effectively.

**Step 6: Profiling Segments**

**Identifying Key Characteristics of Market Segments:**

The aim of the profiling step is to get to know the market segments resulting from the extraction steps. Profiling is only required when data-driven market segmentation is used. For common-sense segmentation, the profiles of the segments are predefined.

If, for example, age is used as the segmentation variable for the commonsense segmentation, it is obvious that the resulting segments will be age groups. At the profiling stage, we inspect a few alternative market segmentation solutions. This is particularly important if no natural segments exist in the data, and either a reproducible or a constructive market segmentation approach must be taken.

**Traditional Approaches to Profiling Market Segments:**

Data-driven segmentation solutions are usually presented to users (clients, managers) in one of two ways: (1) as high-level summaries simplifying segment characteristics to a point where they are misleadingly trivial, or (2) as large tables that provide, for each segment, exact percentages for each segmentation variable.

Such tables are hard to interpret, and it is virtually impossible to get a quick overview of the key insights.

Using Table as the basis of interpreting segments shows that the defining characteristics of segment 2, for example, are: being motivated by rest and relaxation, and not wanting to exceed the planned travel budget. Also, many members of segment 2 care about a change of surroundings, but not about cultural offers, an intense experience of nature, about not caring about prices, health and beauty and realizing creativity. Segment 1 is likely to be a response style segment because – for each travel motive – the percentage of segment members indicating that a travel motive is relevant to them is low (compared to the overall percentage of agreement).

Sometimes – to deal with the size of this task – information is provided about the statistical significance of the difference between segments for each of the segmentation variables. This approach, however, is not statistically correct. Segment membership is directly derived from the segmentation variables, and segments are created in a way that makes them maximally different, thus not allowing to use standard statistical tests to assess the significance of differences.

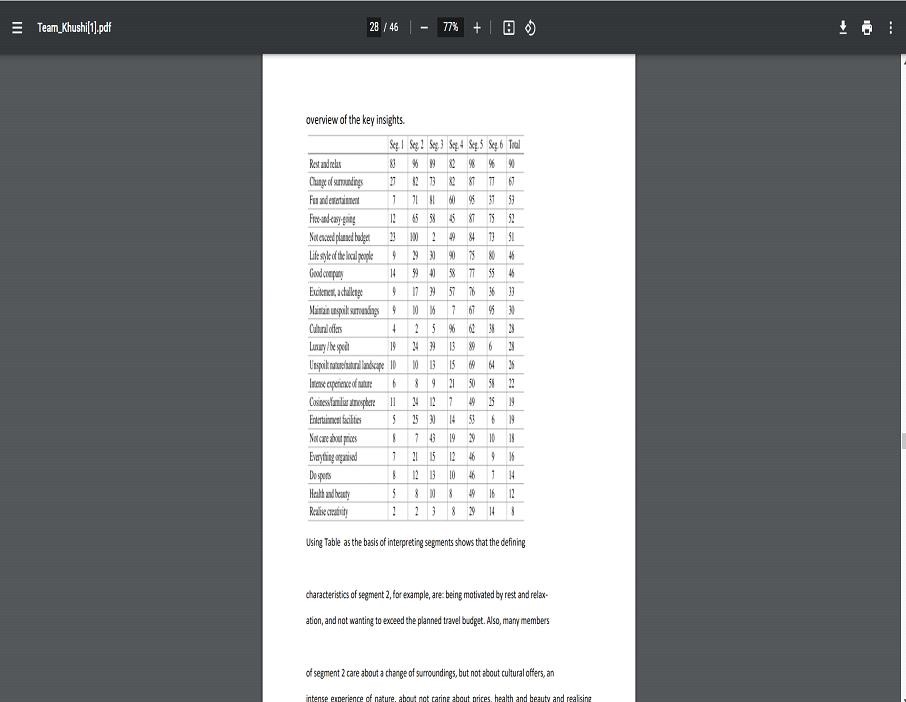


**Segment Profiling with Visualizations:**

Neither the highly simplified, nor the very complex tabular representation typically used to present market segmentation solutions make much use of graphics, although data visualisation using graphics is an integral part of statistical data analysis Graphics are particularly important in exploratory statistical analysis (like cluster analysis) because they provide insights into the complex relationships between variables. In addition, in times of big and increasingly bigger data, visualization offers a simple way of monitoring developments over time. Visualizations are useful in the data-driven market segmentation process to inspect, for each segmentation solution, one or more segments in detail. Statistical graphs facilitate the interpretation of segment profiles. They also make it easier to assess the usefulness of a market segmentation solution. The process of segmenting data always leads to many alternative solutions. Selecting one of the possible solutions is a critical decision. Visualisations of solutions assist the data analyst and user with this task.

**Identifying Defining Characteristics of Market Segments:**

A good way to understand the defining characteristics of each segment is to produce a segment profile plot. The segment profile plot shows – for all segmentation variables – how each market segment differs from the overall sample. The segment profile plot is the direct visual translation of tables such as Table 8.1. The t() around the data matrix vacmot transposes the matrix such that distances between columns rather than rows are computed. Next, hierarchical clustering of the variables is conducted using Ward’s method. Figure 8.1 shows the result. Argument which specifies the variables to be included, and their order of presentation. Here, all variables are shown in the order suggested by hierarchical clustering of variables. shade = TRUE identifies so-called marker variables and depicts them in color. These variables are particularly characteristic for a segment.



**Step 7: Describing Segments**

**7.1 Developing a Complete Picture of Market Segments:**

1**.Segment Profiling:** Segment profiling involves understanding the differences in segmentation variables across market segments. These variables are chosen during the early stages of market segmentation analysis, both conceptually and empirically. They form the basis for extracting market segments from data. The goal of segment profiling is to investigate the variations in segmentation variables among different segments and gain a deeper understanding of their characteristics.

**2.Segment Description:** Segment description, which is like segment profiling, focuses on providing additional information about market segments using variables that were not used for extracting the segments. This additional information can include demographic, psychographic, socio-economic variables, media exposure, and specific product and brand attitudes or evaluations. Describing market segments helps to create a comprehensive picture of each segment and facilitates the development of a customized marketing mix.

**3.Importance of Market Segment Descriptions:** Detailed market segment descriptions are crucial for gaining insights into the nature of segments and developing a customized marketing strategy. By understanding the unique characteristics of each segment, marketers can tailor their messaging, communication channels, and offerings to effectively reach and engage with the target audience. Descriptions of market segments also help identify specific opportunities and challenges associated with each segment, allowing for more informed decision-making.

**7.2 Using Vizualisations to Describe Market Segments:** Visualizations play a vital role in describing market segments as they simplify the interpretation of results for both analysts and users. They provide a graphical representation of differences in descriptor variables and can integrate information on the statistical significance of these differences. Visualizations not only enhance the understanding of market segment characteristics but also help avoid over-interpretation of insignificant differences. Graphical displays are often preferred by marketing managers for their intuitive and efficient presentation of research results.

**7.2.1 Nominal and Ordinal Descriptor Variables:**

When describing differences between market segments in nominal or ordinal descriptor variables (such as gender or level of education), crosstabulation is a commonly used method. This involves creating a table that shows the distribution of segment membership across the descriptor variable categories. Visualizations, such as stacked bar charts or mosaic plots, can be used to represent these crosstabulations. They provide a clear view of segment sizes and allow for easy comparison of proportions across segments.

**Benefits of Visualizing Nominal and Ordinal Descriptor Variables:**

Visualizing nominal and ordinal descriptor variables help simplify the interpretation of results and enables a quick understanding of segment differences. It allows marketers to identify any distinct patterns or variations in segment characteristics, such as gender distribution, educational backgrounds, or country of origin. By visually representing these differences, marketers can make informed decisions about targeting specific segments and developing tailored marketing strategies to effectively engage with each segment's unique characteristics.

1. Gender Distribution Across Market Segments
2. Income Variation Among Market Segments
3. Association Between Travel Motives and Environmental Obligation
   1. **Testing for Segment Differences in Descriptor Variables:**

Statistical tests can be used to formally test for differences in descriptor variables across market segments. The chi-square test can be used to test for independence between a nominal segment membership variable and another nominal or ordinal variable. The results of the test can be visualized using a mosaic plot. For metric variables, such as age or dollars spent on accommodation, analysis of variance (ANOVA) is commonly used to test for significant differences in means between multiple market segments. The F-test is performed, and if the p-value is smaller than 0.05, it indicates that at least two segments differ in their means. Pairwise comparisons between segments can provide more detailed information about which segments differ significantly from each other. Tukey's honest significant differences test can be used to visualize and interpret these pairwise comparisons. Additionally, p-values should be adjusted for multiple testing to control the overall error rate. Bonferroni correction and methods such as Holm's procedure or the false discovery rate procedure can be used for this purpose.

* 1. **Predicting Segments from Descriptor Variables:**

In this approach, we use regression models to predict market segment membership based on descriptor variables. The regression model treats segment membership as the categorical dependent variable and the descriptor variables as independent variables. This allows us to simultaneously test the differences in all descriptor variables in relation to segment membership. The prediction performance of the regression model indicates how well we can identify members of a market segment based on the descriptor variables. Additionally, we can determine which descriptor variables are critical in identifying segment membership, especially if variable selection methods are used. The basic regression model is the linear regression model, which assumes that the dependent variable can be predicted using the independent variables. It assumes a linear relationship between the dependent variable and the independent variables and assumes that the dependent variable follows a normal distribution with a mean determined by the independent variables.

In R, the lm() function is commonly used to fit a linear regression model. The formula interface in R allows us to specify the dependent variable and independent variables. In the case of categorical variables, such as segment membership, they are appropriately coded as factors. The text further discusses the interpretation of the intercept in binary logistic regression, which represents the value of the linear predictor when all independent variables are zero. By applying the inverse logit function to the intercept, one can obtain the predicted probability of belonging to a certain segment for a specific combination of predictor values. Additionally, the text mentions model selection techniques to address the inclusion of irrelevant variables and potential overfitting. The "step" function in R is highlighted as an example of a stepwise procedure that evaluates the improvement in model fit based on the AIC criterion and iteratively adds or drops variables to optimize the model's performance. Finally, the text briefly touches on the assessment of predictive performance using the "predict()" function in R. By obtaining predicted probabilities for each observation and comparing the distributions of predicted probabilities for members and non-members of a segment, one can evaluate how well the model distinguishes between the two groups. In summary, the text provides an overview of binary logistic regression in the context of GLMs, explains the interpretation of coefficients, describes the use of the glm() function and the "effects" package in R, mentions model selection techniques, and briefly discusses the assessment of predictive performance. 7.4.2 Multinomial Logistic Regression: Multinomial logistic regression is used when the dependent variable has more than two categories. Overall, multinomial logistic regression allows for the simultaneous prediction of multiple categories and provides insights into the relationship between independent variables and the probability of belonging to each category.

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

**The Targeting Decision:**

The targeting decision is a crucial step in market segmentation where the selection of one or more target segments is made. This decision has a significant impact on the future performance of an organization. After conducting a global market segmentation analysis and profiling various segments, Step 8 involves selecting the market segments for targeting. The selection process takes into account previously established knockout criteria and segment attractiveness criteria. The goal is to identify segments that align with the organization's objectives and have the potential for long-term commitment.

**Market Segment Evaluation:**

To facilitate the evaluation of alternative market segments, the use of a decision matrix is recommended. This matrix visualizes the relative segment attractiveness and organizational competitiveness for each market segment. The criteria used in the matrix cover two dimensions: segment attractiveness and relative organizational competitiveness. The segment attractiveness represents how desirable a segment is to the organization, while organizational competitiveness indicates how likely the organization is to be chosen by the segment. By plotting the segments on the matrix, the segmentation team can assess their overall attractiveness and competitiveness.

**Step 9: Customizing the Marketing Mix**

In today's competitive business landscape, companies must adapt their marketing strategies to meet the unique needs and preferences of their target segments. This report aims to explore the implications of marketing mix decisions and provide insights into customizing the marketing mix for maximum effectiveness. By studying the profile and detailed description of the target segment, we can determine how product related aspects should be designed or modified to best cater to their specific requirements. Additionally, reviewing the marketing mix enables us to align our strategies with the target segment's preferences.

**Implications of Marketing Mix Decisions:**

The marketing mix consists of the 4 P’s: product, price, place, and promotion. Each element plays a crucial role in influencing customer behavior and satisfaction. When customizing the marketing mix, it is essential to consider the following implications:

* **Product:** Analyze the target segment's needs, preferences, and pain points. Identify areas where the existing product can be modified or new products can be developed to address these specific requirements. Customization might involve features, packaging, branding, or even the creation of entirely new product lines.
* **Price:** Evaluate the target segment's price sensitivity and purchasing power. Determine the optimal pricing strategy, which may involve competitive pricing, value-based pricing, or premium pricing. It is crucial to strike a balance between affordability for the target segment and profitability for the company.
* **Place:** Understand the target segment's preferred distribution channels and buying habits. Ensure the product is available where and when the segment is most likely to make a purchase. This may involve exploring different retail channels, e-commerce platforms or even establishing partnerships with relevant businesses.
* **Promotion:** Develop a tailored communication strategy to effectively reach the target segment. Identify the most appropriate marketing channels such as social media, influencer partnerships or traditional advertising, based on the segment's media consumption habits. Craft messages that resonate with their values, interests, and aspirations.



Fig 1.4 P’s of Marketing Matrix

**Profile and Detailed Description of the Target Segment:**

To customize the marketing mix successfully, a comprehensive understanding of the target segment is essential. Conduct market research, surveys, and customer interviews to gather demographic, psychographic, and behavioral data. Analyze the information to identify common characteristics, motivations, and pain points within the target segment. This data-driven approach will help create a detailed profile that forms the foundation for customization.

**Product-Related Aspects for Target Segment:**

Based on the target segment's profile, determine how the product-related aspects should be designed or modified. Consider the following:

⦁ **Features:** Adapt the product features to align with the target segment's needs and preferences. Identify unique selling points that differentiate the product from competitors and cater to the segment's specific requirements.

**Packaging:** Design packaging that appeals to the target segment visually and communicates the product's value proposition effectively. Consider factors such as convenience, eco-friendliness, and brand image in packaging decisions.

⦁ **Branding:** Develop a brand identity that resonates with the target segment's values and aspirations. Craft compelling brand messaging that communicates the product's benefits and connects emotionally with the segment.

**Price-Related Aspects for Target Segment:**

When customizing the marketing mix, it is crucial to consider price related aspects that align with the target segment's preferences and purchasing power. Here are some key considerations:

⦁ **Pricing Strategy:** Determine the most appropriate pricing strategy based on the target segment's price sensitivity and willingness to pay:

1. **Competitive Pricing:** Set prices in line with or slightly below competitors to attract price-conscious customers within the segment.
2. **Value-Based Pricing:** Emphasize the unique value proposition of the product and set prices accordingly. This strategy works well if the target segment perceives the product as offering superior benefits or solving specific pain points.
3. **Premium Pricing:** Position the product as a high-end or luxury option, charging a premium price to cater to a segment that values exclusivity or premium features.

⦁ **Discounts and Promotions:** Evaluate whether offering discounts, promotions, or bundle deals aligns with the target segment's preferences. This can be particularly effective for price-sensitive segments or during specific periods (e.g. seasonal sales).

⦁ **Payment Options:** Consider providing flexible payment options that accommodate the target segment's financial preferences. This may include installment plans, subscription models, or alternative payment methods (e.g. mobile wallets).

**Promotion-Related Aspects for Target Segment:**

Promotion plays a vital role in reaching and engaging the target segment. To customize the marketing mix effectively, consider the following promotion-related aspects:

**Communication Channels:** Identify the most relevant communication channels to reach the target segment. This may include social media platforms, industry-specific publications, online forums, or offline channels such as events or print media.

⦁ **Messaging and Content:** Craft promotional messages and content that resonate with the target segment's values, interests, and aspirations. Use language, imagery, and storytelling techniques that speak directly to their needs and desires. Personalization and customization can enhance the effectiveness of promotional efforts.

⦁ **Influencer Partnerships:** Collaborate with influencers or thought leaders who have influence and credibility within the target segment. Their endorsement or promotion of the product can significantly impact the segment's perception and purchasing decisions.

⦁ **Targeted Advertising:** Utilize targeted advertising strategies, such as demographic and interest-based targeting, to reach the specific individuals within the target segment.

⦁ **Word-of-Mouth Marketing:** Encourage positive word-of-mouth by providing exceptional customer experiences, incentivizing referrals, or implementing loyalty programs.

⦁ **Event Marketing:** Consider organizing or participating in events that are relevant to the target segment's interests or industry. This provides an opportunity to engage directly with potential customers, showcase the product, and build brand awareness.

**Place-Related Aspects for Target Segment:**

Place-related aspects are crucial when customizing the marketing mix to cater to the target segment's preferences and buying habits. Here are key considerations for place-related aspects:

⦁ **Distribution Channels:** Identify the most appropriate distribution channels that align with the target segment's shopping behavior and preferences.

⦁ **Geographic Reach:** Determine the geographical locations where the target segment is concentrated or where there is a high demand for the product.

⦁ **Retail Partnerships:** Establish partnerships with retailers or businesses that have a strong presence among the target segment.

**Online Presence and User Experience:** Invest in an optimized online presence, including user-friendly websites, mobile applications, or dedicated online stores.

⦁ **Supply Chain Management:** Optimize the supply chain to ensure timely and efficient delivery of the product to the target segment.

⦁ **After-Sales Support:** Consider the target segment's need for aftersales support or customer service.

Reviewing the Marketing Mix in its Entirety:

After customizing the product-related aspects, it is essential to review the entire marketing mix to ensure consistency and alignment. Consider the interplay between product, price, place, and promotion and evaluate:

⦁ **Coherence:** Verify that all elements of the marketing mix align with each other and reinforce the desired brand image and positioning for the target segment.

⦁ **Integration:** Ensure that the chosen marketing channels, messaging, and pricing strategy work together seamlessly to create a cohesive customer experience.

⦁ **Measurement:** Establish key performance indicators (KPIs) to track the effectiveness of the marketing mix in reaching and engaging the target segment. Regularly monitor and analyze data to make informed adjustments as needed.