(2ND PROJECT STUDY TASK)
Report on

Market Segmentation

- Understanding It, Doing It, and Making It Useful

Table of Contents

Abstract

The online vehicle booking market has experienced significant growth in recent years due to the proliferation of internet services and the increasing demand for efficient transportation solutions. This abstract provides an overview of the current state of the online vehicle booking industry, highlighting key trends, challenges, and opportunities that shape its landscape.

The online vehicle booking market has transformed the traditional way people access transportation services. Enabled by advanced technology and innovative platforms, consumers can now conveniently book a variety of vehicles ranging from taxis and rideshares to rental cars and chauffeur services. This transformation has been driven by various factors, including the growing preference for hassle-free and on-demand transportation options, the rise of mobile applications facilitating real-time bookings, and the increasing focus on sustainable and eco-friendly transport alternatives.

However, the industry also faces some challenges. The rapid proliferation of online vehicle booking services has led to intense competition among players, creating pricing pressures and a focus on differentiation through service quality and user experience. Additionally, the sector grapples with issues related to safety, security, and regulatory compliance, particularly concerning passenger and driver verification, data privacy, and insurance coverage.

Amidst these challenges, numerous opportunities lie ahead for the online vehicle booking market. The emergence of autonomous and electric vehicles presents a promising avenue for enhancing the sustainability and efficiency of transportation services. Moreover, the integration of artificial intelligence and machine learning into booking platforms can lead to personalized recommendations, optimized route planning, and improved customer support.

The COVID-19 pandemic has also played a crucial role in shaping the industry's trajectory. While it caused a temporary setback with decreased travel demand, it accelerated the adoption of contactless and hygiene-conscious services, creating a permanent shift towards digital bookings and virtual payment methods.

To thrive in this competitive landscape, industry players must continually adapt to changing customer preferences, invest in technology upgrades, and establish robust safety measures. Collaboration with local governments and regulators is essential to address legal and ethical concerns effectively.

Market Analysis

Market segmentation is a decision-making tool used in the process of defining a target market for a specific product and building an acceptable marketing mix. It is important to understand the implications of pursuing a market segmentation strategy. Segmentation criteria are characteristics of consumers that are essential for market segmentation. One specific consumer feature, such as age, gender, place of origin, etc., might serve as the segmentation criterion.

Market segmentation lets companies focus on specific client segments rather than the whole public, which makes planning campaigns easier. Marketers may use time, money, and other resources more effectively by segmenting their customers. Market segmentation enables businesses to understand their clientele. They are able to customize ads to the client groups most likely to make purchases because they have a better grasp of the requirements and wants of their customers.

The decision to investigate the potential of a market segmentation strategy must be made at the highest executive level, and must be systematically and continuously communicated and reinforced at all organizational levels and across all organizational units. Market segmentation forces organizations to take stock of where they stand, and where they want to be in future. When implemented well, it also leads to tangible benefits, including a better understanding of differences between consumers, which improves the match of organizational strengths and consumer needs.

Market Segmentation Analysis(MSA)

The two main parts of any organization's business marketing plan are 1) Data Analysts and 2) User. Data analysts can provide facts about these segments, but cannot select the most suitable ones. This selection is driven, in part, by the strengths and opportunities of the organisation, and their alignment with the key needs of the market segments. Finally, as soon as one or more target segments have been selected, users need to develop a marketing plan for those market segments, and design a customised marketing mix.

Layers

Analysts make report by taking below steps also called as layers of MSA

Layer-1: Conducting high quality market segmentation analysis. Extracting market segments from the customers data.

Layer-2: Enabling high quality market segmentation analysis. Collecting good data, exploring data, profiling segments, describing segments.

Layer-3: Making it happen in practice. Deciding to segment, defining the ideal segment, selecting (the) target segment(s), developing a customised marketing mix, assessing effectiveness and monitoring marketing changes.

MSA step-by-step

Step 01: Deciding (not) to segment

Step 02: Specifying the ideal target segment

Step 03: Collecting Data

Step 04: Exploring Data

Step 05 Extracting Segments

Step 06: Profiling Segments

Step 07: Describing Segments

Step 08: Selecting the target segments

Step 09: Customising the Market mix

Step 10: Evaluation and Monitoring

Step 1: Deciding (not) to Segment

Before investing time and resources in the market segment analysis, it is important that the organization understands the need to commit to the segmentation strategy in the long term and have willingness and ability to make substantial changes and investments. Strategic business units in charge of segments offer a suitable organizational structure to ensure ongoing focus on the needs of market segments. The decision to investigate the potential of a market segmentation strategy must be made at the highest executive level and must be consistently communicated to and reinforced across all organizational levels due to the significant implications of such a long-term organizational commitment.

Implications of committing to Market Segmentation:

- The key implication is that the organisation needs to commit to the segmentation strategy on the long term.
- There are costs for performing research, field surveys, focus groups, designing multiple packages, and designing multiple advertisements and communication messages.
- Potentially required changes include the development of new products, the modification of existing products, changes in pricing and distribution channels used to sell the product, as well as all communications with the market.

Implementation Barriers:

- Senior Management - Poor Leadership, pro-active championing, Commitment and Involvement in Segmentation Process.
- Organisational Culture - Lack of market or customer orientation, No innovative ideal changes, no proper communication, lack of sharing information and insights, short-term thinking, unwillingness to make change and office politics.
- Lack of Training – No proper basics of market segmentation.
- Formal Marketing Function - lack of qualified experts and Analysts worked on Market Segmentation Analysis.
- Organisational Barriers - Resources, Costs of implementation, poor structural changes in organisation etc.
- Process Barriers - lack of objectives, poor planning, structural process to guide the team etc.

Step 2 - Specifying the Ideal Target Segment

After having committed to investigate the value of a segmentation strategy with the help of "Checklist", the organization has to make a major contribution to market segmentation analysis in this stage. While this contribution is conceptual in nature, it guides many of the following steps, most critically data collection and selecting one or more target segments.

Segment Evaluation Criteria:

The Organisation must determine two Evaluation criteria as below:

- **Knock-out Criteria:** These criteria are the essential, non-negotiable features of segments that the organisation would consider targeting. Knock-out criteria are used to determine if market segments resulting from the market segmentation analysis qualify to be assessed using segment attractiveness criteria.
- **Attractiveness Criteria:** These criteria are used to evaluate the relative attractiveness of the remaining market segments – those in compliance with the knock-out criteria. Attractiveness criteria are not binary in nature. Segments are not assessed as either complying or not complying with attractiveness criteria.

Implementing a Structured Process:

The most popular structured approach for evaluating market segments in view of selecting them as target markets is the use of a segment evaluation plot showing segment attractiveness along one axis, and organisational competitiveness on the other axis. The segment attractiveness and organisational competitiveness values are determined by the segmentation team. This is necessary because there is no standard set of criteria that could be used by all organisations.

At the end of this step, the market segmentation team should have a list of approximately six segment attractiveness criteria. Each of these criteria should have a weight attached to it to indicate how important it is to the organisation compared to the other criteria.

Step 3 - Collecting Data

3.1 Segmentation Variable

Segmentation variables, also known as market segmentation variables, are characteristics or attributes used to divide a larger population into distinct groups or segments. These variables help businesses and marketers understand and target specific customer segments with tailored marketing strategies.

3.2 Segmentation Criteria

3.2.1 Geographic segmentation

Geographical segmentation is a market segmentation strategy that divides a target market based on geographic factors or locations. By categorizing consumers or businesses into distinct groups based on their geographic location, such as country, region, state, or city, companies can better understand and cater to the specific needs and preferences of customers in different locations. This approach allows businesses to tailor their marketing strategies, product offerings, and distribution channels to suit the characteristics and demands of each geographic segment, resulting in more targeted and effective marketing campaigns.

3.2.2. Socio-Demographic Segmentation

Socio-demographic segmentation is a marketing strategy that divides a target market based on social and demographic factors. By analyzing characteristics such as age, gender, income, education, occupation, marital status, and other demographic variables, businesses can identify distinct customer segments with similar sociological and demographic profiles. This segmentation helps companies understand the diverse needs, preferences, and behaviors of different customer groups, enabling them to develop tailored marketing strategies, products, and services that cater to the specific demands of each socio-demographic segment.

3.2.3 Psychographic Segmentation

Psychographic segmentation is a marketing approach that divides a target market based on psychological and lifestyle factors. By analyzing individuals' attitudes, values, interests, opinions, personality traits, and social class, businesses can understand their customers on a deeper level. This segmentation allows companies to create marketing strategies that resonate with specific psychographic segments, tailoring products, messages, and communication channels to align with the unique motivations and preferences of each segment. Psychographic segmentation helps businesses connect with customers on an emotional level, building stronger brand loyalty and increasing the effectiveness of their marketing efforts.

3.2.4 Behavioural Segmentation

Behavioral segmentation is a marketing strategy that divides a target market based on consumer behaviors, actions, and interactions with products or services. It involves analyzing factors such as purchasing patterns, brand loyalty, product usage, benefits sought, decision-making processes, and customer loyalty. By understanding the behavioral

characteristics of customers, businesses can develop targeted marketing strategies and tailor their offerings to meet the specific needs and preferences of each segment. Behavioral segmentation helps companies identify key trends, predict future buying behaviors, and effectively position their products or services to drive engagement, loyalty, and ultimately, increase customer satisfaction and business success.

3.3 Data from Survey Studies

3.3.1 Choice of variable

Avoid unnecessary variables because it causes respondent fatigue, increases the dimensionality of the segmentation problem without adding relevant information, and diverts the attention of the segment extraction algorithm away from information critical to the extraction of optimal market segments.

The recommendation is to ask all necessary and unique questions, while resisting the temptation to include unnecessary or redundant questions. This process involving both *qualitative, exploratory and quantitative survey research* ensures that no critically important variables are omitted.

Step 4 - Exploring Data (Khushi)*:*

Data must be cleaned up after collection and, if necessary, pre-processed. Data exploration is the term for this. Finding the variables' measurement levels, examining each variable's univariate distribution, and evaluating the dependency patterns across variables are all helpful. Results from the data exploration step shed light on whether certain segmentation techniques are effective for isolating market segments.

Cleaning the data is the first step before starting data analysis. This entails verifying that all values have been recorded accurately and that the levels of categorical variables have been given consistent labels. Before using data for analysis, it is advisable to examine the range of possible values for metric variables and the permitted category values for categorical variables.

Following data cleaning, we can perform a descriptive analysis to gain a general understanding of the data through descriptive numerical and graphical representations. We may learn some fundamental information about each column in the dataset using the describe () function. Histograms, box plots, and scatter plots are useful graphical techniques for numerical data. Frequency count bar charts are helpful for visualizing categorical data. The relationship of numerous categorical variables is shown through mosaic plots. The distribution of numerical variables is shown via histograms. They display the frequency of

observations falling inside a given value range. Histograms show whether a variable's distribution is symmetrical and unimodal, or if it is skewed. Some significant inferences about the dataset can be made by examining these graphs for the given dataset.

Preprocessing Categorical Variables

For categorical variables, two pre-processing techniques are frequently utilised. Before further analysis, one method combines levels of category variables; the other, if appropriate, converts categorical variables to numeric ones. If the initial categories are very distinct (have too many subcategories), merging levels of categorical variables can be helpful. The measurement level or scale of variables is an assumption made by many data analysis techniques. The categorical variables can occasionally be converted to numeric variables. If it can be assumed that the lengths between neighbouring scale points on the ordinal scale are roughly equal, ordinal data can be transformed into numerical data. It may be preferable to utilize binary answer alternatives unless there is a compelling case for employing multi-category scales (with unclear distances between scale points).

Binary response options don't need data pre-processing and are less likely to capture response styles. The data is unavoidably changed in some way by pre-processing. When there are just two categories, most statistical procedures perform correctly after converting binary data to numeric variables. The logical matrix obtained by comparing the items in the data frame to the string "yes" is converted to a numeric matrix with 0 denoting FALSE and 1 denoting TRUE by adding 0 to the logical matrix.

Preprocessing Numeric Variables

The range of values of a segmentation variable affects its relative influence in distance-based methods of segment extraction. To balance the influence of segmentation variables on segmentation results, variables can be standardized. Standardizing variables means transforming them in a way that puts them on a common scale. The default standardization method in statistics subtracts the empirical mean $\bar{x}$ and divides by the empirical standard deviations:

$$z_i = \frac{x_i - \bar{x}}{s},$$

$$\text{with} \qquad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad\qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2,$$

for the n observations of a variable $x = \{ x_1 , ..., x_n \}$.

Principal Components Analysis:

Finding the most significant characteristics or principal components that have the greatest influence on the target variable is done through a procedure called P**rincipal Components Analysis (PCA)**. It creates a new data set with variables known as principal components that are uncorrelated and ordered by importance from a multivariate data set containing metric variables. The majority of the variability is found in the first variable (principal component), followed by the second principal component and so on. Because principal components analysis creates the same number of new variables as there were old ones, the observations (consumers) maintain their original relative positions to one another after transformation. Principal components analysis essentially maintains the same data space while seeing it from a different perspective.

Prior to creating market segments from consumer data, principal components analysis is sometimes done to reduce the number of segmentation variables. This notion is intriguing because it increases the dimensionality of the problem that the segment extraction technique must handle, making extraction more challenging and necessitating larger sample sizes.

Principal components analysis can be used to investigate data and find strongly associated variables, but it is not advised to use a subset of principal components as segmentation variables. High loadings on the same principal components, a sign of redundancy in the data they capture, are displayed by highly linked variables. Some of the initial, redundant variables in the segmentation base can be eliminated using the insights from such an exploratory study. This method still uses the initial variables gathered, but it also reduces the dimensionality.

Step 5 – Extracting Segments (Himanshu & Shantanu)

Grouping Customers:

It is very important to group consumer data, as data comes in all shapes and forms but there is no single best algorithm for all data sets. If consumer data is well-structured, and well-separated, distinct market segments exist, tendencies of different algorithms matter less. If,

However, data is not well-structured, the tendency of the algorithm influences the solution substantially. In such situations, the algorithm will impose a structure that suits the objective function of the algorithm.

Two methods for segmenting:

1. Distance-Based Method
2. Model-Based Method

Distance-Based Method:

Distance Measure:

A distance measure has to comply with a few criteria:

1. Symmetry

$$d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x}).$$

2. Distance of a vector to itself and only to itself is 0

$$d(\mathbf{x}, \mathbf{y}) = 0 \Leftrightarrow \mathbf{x} = \mathbf{y}.$$

3. Fulfill the so-called triangle inequality

$$d(\mathbf{x}, \mathbf{z}) \leq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z}).$$

Two Mostly used distances in market segmentation:

1. Euclidean Distance (Most Commonly used):

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{j=1}^{p} (x_j - y_j)^2}$$

2. Manhattan or Absolute Distance

$$d(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^{p} |x_j - y_j|$$

3. Asymmetric Binary Distance (only for binary vector)

$$d(\mathbf{x}, \mathbf{y}) = \begin{cases} 0, & \mathbf{x} = \mathbf{y} = \mathbf{0} \\ (\#\{j | x_j = 1 \text{ and } y_j = 1\})/(\#\{j | x_j = 1 \text{ or } y_j = 1\}) \end{cases}$$

Note: Data needs to be standardized before calculating distances

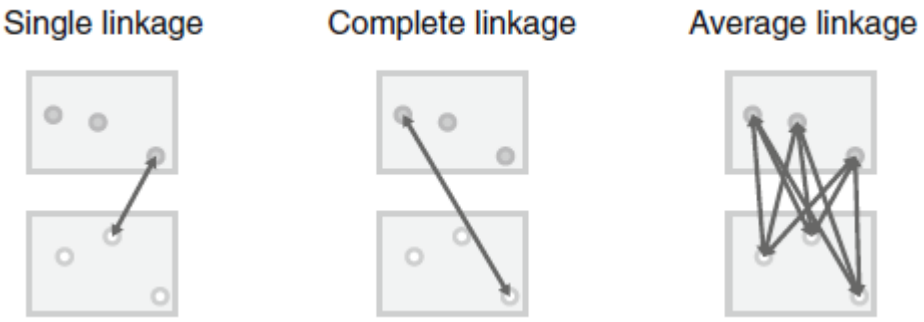
Hierarchical Methods:

Two methods:

1. *Divisive hierarchical clustering* methods start with the complete data set X and splits it into two market segments
2. *Agglomerative hierarchical clustering*: the two market segments closest to one another are merged until the complete data set forms one large market segment.

The *Linkage Method* generalizes how, given a distance between pairs of observations, distances between groups of observations are obtained

1. Single linkage: distance between the two closest observations of the two sets.
2. Complete linkage: distance between the two observations of the two sets that are farthest away from each other.
3. Average linkage: mean distance between observations of the two sets.



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.

Model Based Methods

Model-based methods in market segmentation refer to approaches that utilize statistical models to identify distinct segments or groups within a market. These methods aim to partition consumers or observations into homogeneous segments based on their characteristics, behaviors, preferences, or other relevant factors.

1. *Finite Mixture Model:* A finite mixture of distributions is a statistical model that represents a population as a combination of several component distributions. Each component distribution represents a distinct subgroup or cluster within the population. The finite mixture model assumes that the observed data points are generated from one of these component distributions, and the goal is to estimate the parameters of the component distributions and assign each data point to its most likely cluster.

A finite mixture of distributions can be represented by the following equation:

$$\sum_{h=1}^{k} \pi_h f(y|\theta_h), \quad \pi_h \geq 0, \quad \sum_{h=1}^{k} \pi_h = 1.$$

In short, the equation sums up the weighted contributions of each component distribution to define the overall PDF of the finite mixture model. This allows for capturing complex data patterns and identifying distinct segments or clusters within the population.

2.*Finite Mixture of Regressions:* Finite mixtures of regression models are a statistical approach used to model heterogeneity in regression analysis. They extend the concept of finite mixtures to regression modeling, allowing for the identification of distinct subgroups or clusters with different regression relationships within a population.

In a finite mixture of regression models, the observed data points are assumed to come from a mixture of several component regression models. Each component regression model represents a distinct subgroup or cluster within the population. The goal is to estimate the

parameters of the component regression models and assign each data point to its most likely cluster.

The basic equation for a finite mixture of regression models can be expressed as:

$$ y_i = \sum_{k=1}^{K} \pi_k \cdot f_k(x_i) + \epsilon_i $$

By estimating the parameters of the component regression models and determining the probability of each observation belonging to each cluster, finite mixtures of regression models allow for capturing heterogeneity in regression relationships and identifying different subgroups with distinct regression patterns.

3.*Extensions and Variations:*

1. *Variable selection*: Incorporating techniques to select relevant predictor variables for each cluster.
2. *Multivariate regression*: Modeling multiple response variables simultaneously within the mixture model.
3. *Nonlinear regression:* Allowing for nonlinear relationships between predictors and response variables within clusters.
4. *Robustness and outliers:* Handling outliers and improving robustness through specialized estimation techniques.
5. *Bayesian approach:* Formulating the mixture regression model within a Bayesian framework, incorporating prior knowledge and estimating posterior distributions.

These extensions and variations enhance the flexibility and applicability of finite mixtures of regression models to address specific modeling needs, accommodate diverse data characteristics, and provide more tailored insights into the relationships between predictors and response variables within distinct subgroups.

Algorithms with Integrated Variable Selection

Algorithms with integrated variable selection refer to clustering algorithms that automatically perform variable selection as part of the clustering process. These algorithms aim to identify the most relevant subset of variables that contribute significantly to the clustering structure and discard irrelevant or redundant variables. By integrating variable selection within the clustering algorithm, these methods can improve the quality of the clustering results and enhance interpretability.

1.*Biclustering Algorithms:* Biclustering algorithms are techniques used to simultaneously cluster rows and columns of a data matrix, allowing for the identification of subgroups that exhibit similar patterns across both dimensions. Biclustering algorithms are particularly

useful for analyzing datasets where subsets of rows and columns exhibit specific relationships or patterns.

There are several biclustering algorithms available, each with its own approach and characteristics. Here are a few commonly used biclustering algorithms:

1. *K-means Biclustering:* This algorithm applies the K-means clustering algorithm to both rows and columns of the data matrix, seeking to find subsets of rows and columns that form coherent biclusters. It aims to minimize the within-cluster sum of squares for both rows and columns simultaneously.

2. *Spectral Biclustering:* This algorithm utilizes spectral techniques to identify biclusters in the data matrix. It involves eigenvalue decomposition or singular value decomposition to extract the underlying structure and identify coherent biclusters.

3. *Plaid Model:* The Plaid model is based on Boolean matrix factorization. It decomposes the data matrix into binary matrices representing biclusters. The algorithm seeks to find biclusters that exhibit a binary checkerboard pattern, where only specific subsets of rows and columns are active.

4. *Bimax*: Bimax is a biclustering algorithm based on the maximum co-occurrence principle. It identifies biclusters by maximizing the number of co-occurring entries in the data matrix. Bimax aims to find biclusters with dense submatrices that share similar patterns.

5. *Bayesian Biclustering*: Bayesian biclustering algorithms utilize Bayesian modeling techniques to identify biclusters in the data matrix. These algorithms often incorporate prior knowledge, regularization, and Bayesian inference to estimate biclusters.

Each biclustering algorithm has its own strengths and assumptions, and the choice of algorithm depends on the specific characteristics of the dataset and the goals of the analysis. Biclustering algorithms help uncover meaningful structures in complex data by simultaneously clustering rows and columns, enabling researchers to identify subsets of data that exhibit coherent patterns across both dimensions.

2.Variable Selection Procedure for Clustering Binary Data: When dealing with binary data in clustering, variable selection procedures aim to identify a subset of variables that are most relevant for clustering analysis. Here is a general variable selection procedure for clustering binary data:

1. *Calculate Variable Importance:* Calculate a measure of variable importance for each binary variable. Common measures include information gain, chi-square test statistics, or mutual information. These measures capture the discriminatory power of each variable in distinguishing different clusters.

2. *Set a Threshold:* Set a threshold or criteria to determine the importance level for variable selection. You can choose a fixed threshold or use a data-driven approach, such as selecting the top N% of variables based on their importance scores.

3. *Select Relevant Variables:* Select variables that meet the threshold criterion and are deemed relevant for clustering. These variables will form the subset used for clustering analysis.

4. *Cluster the Selected Variables:* Apply a clustering algorithm, such as K-means or hierarchical clustering, using only the selected relevant variables. The algorithm will group observations based on the patterns and similarities in the selected variables.

5. *Evaluate Clustering Results:* Evaluate the quality and interpretability of the clustering results obtained with the selected variables. Assess the within-cluster homogeneity and between-cluster heterogeneity to ensure meaningful and distinct clusters.

6. *Refine Variable Selection:* If the clustering results are not satisfactory, consider refining the variable selection process. This could involve adjusting the threshold, considering alternative measures of variable importance, or exploring other variable selection techniques such as regularization or dimensionality reduction methods.

By following this variable selection procedure, you can identify the subset of binary variables that are most informative for clustering analysis. This helps reduce dimensionality, improve interpretability, and enhance the clustering performance on binary data.

3.*Variable Reduction:* Factor-Cluster Analysis Variable reduction through factor-cluster analysis is a technique that combines factor analysis and cluster analysis to simultaneously reduce the dimensionality of the data and identify meaningful clusters. This approach aims to uncover latent factors that explain the correlations among the variables and then cluster the observations based on these underlying factors.

The general procedure for factor-cluster analysis is as follows:

1. *Factor Analysis:* Conduct factor analysis on the original set of variables to identify the latent factors that explain the common variance among the variables. The number of factors to extract is determined based on eigenvalues, scree plots, or other criteria.

2. *Factor Scores:* Compute factor scores for each observation based on the factor loadings obtained from factor analysis. Factor scores represent the values of each observation on the identified latent factors.

3. *Cluster Analysis:* Perform cluster analysis on the factor scores using a suitable clustering algorithm, such as K-means or hierarchical clustering. This step groups the observations based on their similarities in the derived factor space.

4. *Interpretation:* Examine the clustering results and interpret the clusters based on the original variables and the identified latent factors. This involves analyzing the patterns of the original variables within each cluster and understanding the underlying factors that contribute to the observed cluster structure.

The combined factor-cluster analysis approach provides several advantages. It reduces the dimensionality of the data by replacing the original variables with a smaller set of factor scores. These scores capture the most important information from the original variables, making the analysis more manageable and interpretable. Additionally, it uncovers the latent factors that drive the correlations among the variables, leading to a deeper understanding of the underlying structure of the data.

Factor-cluster analysis is a useful technique for exploratory data analysis, pattern recognition, and discovering hidden relationships within complex datasets. It can be applied in various fields, including social sciences, market research, and customer segmentation, to gain insights into the underlying factors and clusters within the data.

Data Structure Analysis

Market segmentation is an exploratory process regardless of the algorithm used for extraction. Traditional validation methods with clear optimality criteria are not feasible due to practical constraints. Therefore, validation in market segmentation typically refers to assessing the reliability or stability of solutions across repeated calculations, often by modifying the data or algorithm. This stability-based data structure analysis helps identify if natural, distinct, and well-separated market segments exist in the data. If such segments exist, they can be easily revealed. However, if they don't, analysts must explore numerous alternative solutions to identify the most useful segment(s) for the organization. Data structure analysis also assists in determining an appropriate number of segments to extract, considering the presence of clusters or other forms of structure within the data.

Cluster Indices:

Internal cluster indices focus on the compactness and separation of market segments within a single solution. They calculate measures such as the sum of distances between segment members and their representative. The scree plot and Ball-Hall index are examples of internal cluster indices.

External cluster indices, on the other hand, require additional information and compare multiple segmentation solutions. They evaluate the similarity between two solutions,

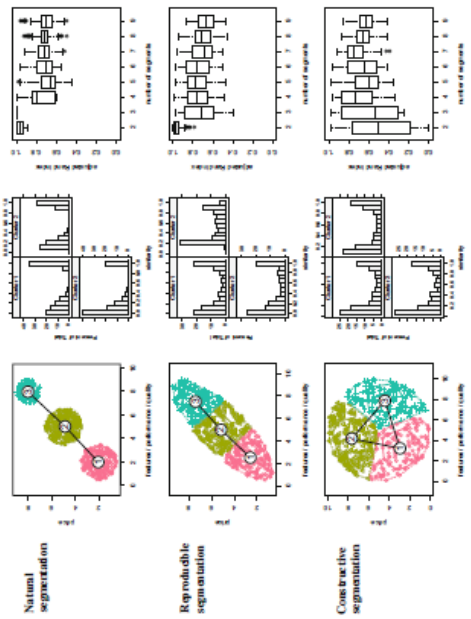
often using measures like the Jaccard index and the Rand index. These indices assess the agreement between pairs of consumers assigned to the same or different segments.

The issue of label switching, where the arbitrary labels of segments can affect the comparison. To address this, the Jaccard index and the Rand index consider pairs of consumers without relying on specific segment labels. The adjusted Rand index, which incorporates a correction for chance agreement based on segment sizes, is also discussed.

Overall, these cluster indices provide insights into different aspects of market segmentation solutions and assist data analysts in making critical decisions, such as selecting the number of market segments to extrac*t.*

Gorge Plots:

A simple method to assess the separation of segments is by examining the distances between each consumer and the segment representatives. The similarity of a consumer to a segment representative can be calculated based on these distances, using a hyperparameter γ that determines the translation of distance differences into similarity differences. These similarity values, ranging from 0 to 1, can be visualized using gorge plots, silhouette plots, or shadow plots. High similarity values indicate close proximity or high probability of segment membership, while low values indicate greater distance or lower probability. This approach is useful for evaluating the quality of segment extraction in both partitioning and model-based clustering methods.



Global Stability Analysis:

An alternative approach to data structure analysis in market segmentation is using resampling methods. These methods assess the stability of segmentation solutions by

generating new data sets and extracting multiple segmentations. The similarity between these solutions is compared to determine the most replicable solution. Resampling methods are particularly useful when the data lacks distinct natural segments. They provide insights into the structure of the data and help determine the most suitable number of segments to extract. The stability analysis can indicate whether natural segments, reproducible segments, or artificially constructed segments exist in the data. This approach is valuable in cases where the data structure cannot be easily determined through visual inspection.

Step 6: Profiling Segments (Harsha)

1. Identifying Key Characteristics of Market Segments

The main goal of this step is to be familiar with the market segments that are derived from extraction. **Profiling** is required only when data-driven market segmentation is used whereas for commonsense segmentation, the profiles of the segments are predefined. The users may think of extracting segments on the basis of benefits sought by consumers.

Even after the data has been analyzed, the foremost characteristics of the resulting market segments are unknown. *Identifying these defining characteristics of market segments with respect to the segmentation variables is the aim of profiling. Profiling not only consists of characterizing the market segments individually, but also in comparison to the other market segments.*

At this stage, we examine a number of 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 has to be taken. Good profiling is the basis for correct interpretation of the resulting segments which in turn is critical to making good strategic marketing decisions.

But the main thing is - *Data-driven market segmentation solutions are not easy to interpret. Most of the studies state that they have difficulties understanding data-driven market segmentation solutions and feel that segmentation analysis is like a black box.*

2. Traditional Approaches for Profiling Market Segments

Data-driven segmentation solutions are usually presented to users in one of the two ways: (1) High level summaries that simplify segment characteristics to an extent where they are misleadingly trivial or (2) Large tables for each segment that provide exact percentages for each segmentation variable. But these tables are hard to interpret and it is virtually

impossible to get a quick overview of the key insights. This is an outrageously tedious task to perform, even for the most astute user.

In some cases, to deal with the size of a 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.

3. Segment Profiling with Visualizations

Although data visualization with graphics is a vital aspect of statistical data analysis, neither the highly simplified nor the very sophisticated tabular form commonly used to demonstrate market segmentation solutions make significant use of visuals.

In exploratory statistical analysis, graphics play a crucial role because they provide light on the intricate interactions between variables. In addition, visualization provides a straightforward method of tracking changes over time in an era of big and getting bigger and greater data.

In the data-driven market segmentation process, visualizations are helpful for closely inspecting one or more segments for each segmentation solution. Segment profile understanding is aided by statistical graphics. Additionally, they make it simpler to evaluate the value of a market segmentation strategy. There are countless potential solutions that can be found when segmenting data. Making a choice among the potential solutions is crucial. The data analyst and user are helped by visualizations of solutions in this process.

3.1. Identifying defining characteristics of Market Segments

Producing a segment profile plot is a wonderful technique to comprehend the distinctive features of each segment. The segment profile plot demonstrates how each market segment varies from the entire sample for all segmentation variables.

The segment profile plot is a type of plot known as a **Panel Plot**. The six panels each depict one portion. The segment profile plot displays the cluster centers (centroids, or representatives of the segments) for each segment.

Managers that rely their long-term strategic decisions on segmentation results can interpret data more easily with the help of good visualizations. Such long-term strategic choices require significant financial investments in the segmentation strategy's execution. Therefore, effective visualizations provide a fantastic return on investment.

3.2. Assessing Segment Separation

A segment separation plot can be used to visualize segment separation. The segment separation plot shows the overlap of segments for all pertinent dimensions of the data space. When there are few segmentation variables, segment separation plots are extremely straightforward; but, as the number of segmentation factors rises, they become more complicated. Segment separation plots, however, provide data analysts and consumers with a fast overview of the data condition and the segmentation solution even in such complex settings.

Step 7: Describing Segments (Harsha)

1. Developing a Complete Picture of Market Segments

Understanding the variations in segmentation factors across market segments is the goal of segment profiling. The selection of segmentation variables occurs **conceptually** in *Step 2 (definition of the ideal target segment)* and **empirically** in *Step 3 (data collection)*. The foundation for deriving market segments from empirical data are segmentation variables.

Comparable to the profiling stage is **Step 7 (describing segments)**. The sole distinction is that market segmentation *has not* been done using the characteristics under examination. Instead, additional information about segment participants is used to describe market segments in Step 7. *Good descriptions of market segments are critical to gaining detailed insight into the nature of segments.*

Profiling, for instance, entails examining distinctions across segments with regard to the travel motivations themselves while doing a data-driven market segmentation analysis utilizing the Australian travel motives data set. Additional data, such as segment members' age, gender, historical travel habits, favourite vacation activities, media consumption, use of information sources during vacation planning, or their spending habits while on vacation, are used in the segment description stage. **Descriptor variables** are the name given to these extra variables.

Inferential statistics can be used to evaluate data, or descriptive statistics, including visualizations, can be used to study differences between market segments with regard to descriptor variables. The marketing literature has historically incorporated statistical analysis and tabular displays of variations in descriptor variables. Segment descriptions are easier to understand when they are visualized.

2. Using Visualizations to Describe Market Segments

There are many different kinds of charts that exist to visualize the difference in descriptor variables. We will go through two basic approaches that are suitable for **nominal** and **ordinal descriptor variables** or **metric descriptor variables.** Two major benefits of using graphical statistics to describe market segments are that: *1) it makes it easier for users and data analysts to understand results* and *2) it incorporates information on the statistical significance of differences, preventing the over-interpretation of insignificant differences.*

2.1 Nominal and Ordinal Descriptor Variables

A cross-tabulation of segment membership with the descriptor variable serves as the foundation for all visualizations and statistical tests when differentiating between market segments using a single nominal or ordinal descriptor variable.

3. Testing for Segment Differences in Descriptor Variables:

Formally testing for variations in descriptor variables across market groups can be done using straightforward statistical tests. Running a set of separate tests for each relevant variable is the simplest way to look for differences. Segment membership, or the assignment of each consumer to a certain market segment, is the result of the segment extraction stage. Segment membership is a nominal variable that can be handled just like any other. It serves as the segmentation variables' nominal summary statistic. Therefore, any test to determine if a nominal variable is associated with another variable is appropriate.

The cross-tabulation of both variables is used as the foundation for the mosaic plot to show the relationship between the nominal segment membership variable and another nominal or ordinal variable. The 2-test is the proper test to determine whether columns and rows in a table are independent.

Analysis of Variance (ANOVA) is the most often used technique for determining whether there are significant differences between the means of more than two groups. We initially examine segment means before testing for variations in mean moral obligation values to safeguard the environment across market segments.

4. Predicting Segments from Descriptor Variables:

A different strategy for understanding market segmentation is to make predictions about segment membership based on descriptor variables. To do this, we employ a regression model in which the independent variables are descriptor variables, and the categorical

dependent variable is segment membership. For classification, we can utilize techniques from statistics, and for supervised learning, from machine learning.

These strategies examine variations in all descriptor variables simultaneously as compared to the methods in the previous section. According to the descriptor variables, the prediction performance shows how well members of a market segment may be identified. Additionally, we discover which descriptor variables—especially if methods for simultaneous variable selection are employed—are essential for determining segment membership.

5. Checklist

Finally, we will make a checklist to be updated with whatever tasks are completed and if there are any pending to make sure we didn't miss any steps.

Step 8 - Selecting the Target Segments (Srinivas)

8.1 Targeting Decision:

At Step-2, knock-out criteria for market segments have been agreed upon, and segment attractiveness criteria have been selected, and weighed to reflect the relative importance of each of the criteria to the organisation. The segmentation team can build on the outcome of Step-2. We have studied that at Step 5, a number of segments are available for detailed inspection. These segments are profiled in Step-6, and described in Step-7. Now, In Step-8 one or more of those market segments need to be selected for targeting.

The first task in Step-8 is to ensure that all the market segments that are still under consideration to be selected as target markets have well and truly passed the knock-out criteria test. Once this is done, the attractiveness of the remaining segments and the relative organisational competitiveness for these segments needs to be evaluated. The below two questions will form the basis of target segment decision:

- Which of the market segments would the organisation most like to target? Which segment would the organisation like to commit to?
- Which of the organisations offering the same product would each of the segments most like to buy from? How likely is it that our organisation would be chosen? How likely is it that each segment would commit to us?

8.2 Market Segment Evaluation:

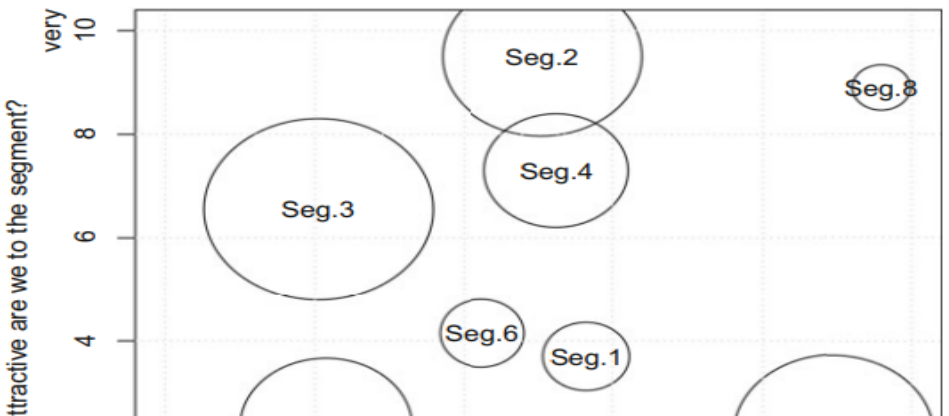
Decision matrix is used to visualise relative segment attractiveness and relative organisational competitiveness for each market segment. There were many such matrices defined in the past to evaluate each market segment. The aim of all these decision matrices along with their visualisations is to make it easier for the organisation to evaluate alternative market segments, and select one or a small number for targeting. It is up to the market segmentation team to decide which variation of the decision matrix offers the most useful framework to assist with decision making.

To keep segment evaluation as intuitive as possible, we plot segment attractiveness along the x-axis, and relative organisational competitiveness along the y-axis. Segments appear as circles. The size of the circles reflects another criterion of choice that is relevant to segment selection, such as contribution to turnover or loyalty.

The target segment selection is the actual value each market segment has for each of the criteria specified to constitute segment attractiveness. These values emerge from the grouping, profiling, and description of each market segment. To determine the attractiveness value to be used in the segment evaluation plot for each segment, the segmentation team needs to assign a value for each attractiveness criterion to each segment.

The location of each market segment in the segment evaluation plot is then computed by multiplying the weight of the segment attractiveness criterion with the value of the segment attractiveness criterion for each market segment. The value of the segment attractiveness criterion for each market segment is determined by the market segmentation team based on the profiles and descriptions resulting from Steps 6 and 7. The result is a weighted value for each segment attractiveness criterion for each segment. Those values are added up, and represent a segment's overall attractiveness. Same procedure is followed for relative organisational competitiveness.

The last aspect of the plot is the bubble size. Anything can be plotted onto the bubble size. Typically profit potential is plotted. Profit combines information about the size of the segment with spending and, as such, represents a critical value when target segments are selected. In other contexts, entirely different criteria may matter. Now the plot is complete and serves as a useful basis for discussions in the segmentation team.



From the above plot, we can observe that segment 5 is obviously highly attractive and has high profit potential, but unfortunately the segment is not as fond of the organisation as the organisation is of the segment. Segment 8 is excellent because it is highly attractive to the organisation, and views the organisation's offer as highly attractive. A match made in heaven, except for the fact that the profit potential is not very high. It may be necessary, therefore to consider including segment 2. Segment 2 loves the organisation, has decent profit potential, and is about equally attractive to the organisation as segments 1, 4 and 6 (all of which, unfortunately, are not very fond of the organisation's offer).

Python code for Segment Evaluation Plot is as below:

```
pip install pyMSA
import matplotlib.pyplot as plt
import numpy as np
from pyMSA import decisionMatrix
size = 100
result = decisionMatrix.decisionMatrix(x, y, wx, wy, size=size)
segments = result['segments']
scores = result['scores']
print(segments)
print(scores)
x_coords = [seg[0] for seg in segments]
y_coords = [seg[1] for seg in segments]
bubble_sizes = np.sqrt(scores) * 50
plt.scatter(x_coords, y_coords, s=bubble_sizes, alpha=0.7)
plt.xlabel('X')
plt.ylabel('Y')
plt.title('Market Segmentation')
plt.show()
```

Where,

x - This represents the x-values or features of the data for segmentation analysis.

y - This represents the y-values or target variable of the data for segmentation analysis.

wx - This is the weight for the x-values. It can be used to assign different weights to different features during the analysis.

wy - This is the weight for the y-values or target variable. It can be used to assign different weights to the target variable during the analysis.

size - This is the size parameter we defined earlier, which determines the size of the decision matrix.

With above plot, Segmentation team will have discussion on Marketing mix which is discussed in next step-9.

STEP 9: Customizing the Marketing Mix (Himanshu)

Implications for Marketing Mix Decisions:

A toolbox to assist in selling products.

Marketing mix consists of the 4Ps: **Product, Price, Promotion and Place**

Segmentation process is frequently seen as part of what is referred to as the segmentation-targeting-positioning (STP) approach

How the target segment decision affects marketing mix development

It is important, however, not to adhere too strictly to the sequential nature of the segmentation-targeting-positioning process. To best ensure maximizing on the benefits of a market segmentation strategy, it is important to customize the marketing mix to the target segment

Product

Purpose is to specify the product in view of customer needs. Often this does not imply designing an entirely new product, but rather modifying an existing one. Marketing mix decisions that fall under the product dimension are: naming the product, packaging it, offering or not offering warranties, and after sales support services.

Price

Typical decisions an organization needs to make when developing the price dimension of the marketing mix include setting the price for a product, and deciding on discounts to be

offered. By segmentation, consumers can be segmented into high and low expenditure groups. Offer discounts in costlier group and to attract from people from other groups.

Place

Purpose is to distribute the product to the customers. This includes answering questions such as: should the product be made available for purchase online or offline only or both; should the manufacturer sell directly to customers; or should a wholesaler or are tailer or both be used. By segmentation, consumers can be segmented into different groups. By considering the median of the box plot of those groups we can place the product in that category.

Promotion

Developing an advertising message that will resonate with the target market, and identifying the most effective way of communicating this message. Other tools in the promotion category of the marketing mix include public relations, personal selling, and sponsorship. By segmentation, consumers can be segmented into different groups. By considering the median of the box plot of those groups we can place the product in that category.

GitHub Links: