Market Segmentation

Assign	
Date	

Step1: Deciding (not) to Segment

▼ Implications of Committing to Market Segmentation

Market segmentation is a marketing strategy where a company divides a larger market into smaller groups with specific needs, preferences, or behaviors. This allows the company to create and market products and services that are tailored to each group. However, pursuing market segmentation requires a long-term commitment and substantial investment in research, product development, marketing, and organizational restructuring. Therefore, a company should only pursue market segmentation if it can generate enough additional sales to justify the cost. It also requires the company to be organized around market segments, rather than products. **The decision to pursue market segmentation must be made at the highest executive level and communicated across the entire organization.**

▼ Implementation Barriers

- Market segmentation is a strategy for dividing a market into smaller groups with distinct needs and characteristics.
- Successful implementation of market segmentation can be impeded by various barriers, including lack of leadership and commitment from senior management, lack of resources, and a culture that is resistant to change and new ideas.
- Lack of training and a formal marketing function or qualified marketing expert can also hinder successful implementation.
- Other obstacles may include objective restrictions, process-related barriers, and lack of understanding of market segmentation techniques.
- Barriers can be identified and proactively removed, but if they cannot be overcome, abandoning the attempt at market segmentation may be necessary.

- Implementing market segmentation requires dedication, patience, and a willingness to face inevitable problems.

Step2: Specifying the Ideal Target Segment

▼ Segment Evaluation Criteria

The third layer of market segmentation analysis depends primarily on user input. The user should be involved in most stages of market segmentation analysis. The organization must determine **two sets of segment evaluation** criteria in step 2 of the process, knock-out criteria and attractiveness criteria, which are used to e**valuate the relative attractiveness of the remaining market segments**. The literature suggests a wide range of possible segment evaluation criteria, which are described in Table 4.1 in the chronological order in which they were proposed.

▼ Knock-Out Criteria

- Knock-out criteria are used to determine if market segments are viable for further analysis.
- The first set of criteria include substantiality, measurability, and accessibility.
- Additional knock-out criteria include segment homogeneity, distinctiveness, size, alignment with organizational strengths, identifiability, and reachability.
- These criteria help ensure that a segment is worth the investment of resources for customized marketing.
- Senior management, the segmentation team, and the advisory committee should understand these criteria, with some criteria needing specific minimum specifications.

▼ Attractiveness Criteria

- Table 4.1 provides a wide range of attractiveness criteria that the segmentation team can consider in step 2 of the process.
- Attractiveness criteria are used to evaluate the relative attractiveness of the remaining market segments.

- These criteria are not binary and do not classify segments as either complying or not complying.
- Each segment is rated based on its attractiveness with respect to a specific criterion.
- The combined attractiveness across all criteria determines whether a market segment is selected as a target segment in step 8 of market segmentation analysis.

▼ Implementing a Structured Process

- Following a structured process when assessing market segments is beneficial.
- The most popular structured approach is the use of a segment evaluation plot.
- The plot shows segment attractiveness along one axis and organizational competitiveness on the other.
- Both segment attractiveness and organizational competitiveness values are determined by the segmentation team.
- A large number of possible criteria has to be investigated before agreement is reached on which criteria are most important.
- A team of people should complete this task.
- The advisory committee, consisting of representatives from all organizational units, should discuss and possibly modify the initial proposal.
- The market segmentation team should have a list of approximately six segment attractiveness criteria, each with a weight attached to it.
- The typical approach to weighting is to ask all team members to distribute 100 points across the criteria.
- Approval by the advisory committee should be sought.

Step3: Collecting Data

▼ Segmentation Variables

The concept being explained is market segmentation, which is the process of dividing a heterogeneous market into smaller, homogeneous subgroups that can be targeted more efficiently and effectively with tailored marketing strategies. Segmentation variables are used to identify or create market segments, and they are based on empirical data. **In commonsense segmentation, the segmentation variable is typically one single characteristic of the consumers** in the sample, while in data-driven segmentation, multiple segmentation variables are used. Descriptor variables are used to describe the segments in detail, and they typically include socio-demographics and information about media behavior.

In the context of market segmentation, the tables are used to illustrate how segmentation variables and descriptor variables are used to create market segments.

Table 5.1 shows an example of commonsense market segmentation where gender is the segmentation variable. Each row in the table represents one consumer, and each variable represents one characteristic of that consumer**. An entry of 1 indicates that the consumer has that characteristic, while an entry of 0 indicates that the consumer does not have that characteristic.** Using gender as the segmentation variable, the sample is split into a segment of women and a segment of men.

Table 5.2, on the other hand, shows an example of data-driven market segmentation, where multiple segmentation variables are used to identify naturally existing or artificially created market segments. In this case, the segmentation variables are the benefits sought when going on vacation, while the socio-demographic variables (gender, age, and number of vacations taken) serve as descriptor variables. The table shows how the sample can be sorted according to these variables to reveal market segments that share a common set of benefits sought when going on vacation.

▼ Segmentation Criteria

Segmentation criteria are the nature of information used for market segmentation.
• It is a crucial decision that must be made by the organization prior to data collection and extraction of segments.
• Segmentation criteria can be a specific construct such as benefits sought or a

broader category such as geographic, socio-demographic, psychographic, or behavioral.
• The decision of which segmentation criterion to use requires prior knowledge about the market and cannot be easily outsourced.
• The most relevant differences between consumers for market segmentation are profitability, bargaining power, preferences for benefits or products, barriers to choice, and consumer interaction effects.
• There are few guidelines for the most appropriate segmentation criteria to use in a given marketing context, but the recommendation is to use the simplest possible approach.
• Using demographic or geographic segmentation may work better and cost less than using more sophisticated segmentation methods such as psychographic segmentation.
• The key is to use the segmentation criterion that works best for the product or service at the least possible cost.

▼ Geographic Segmentation

Geographic segmentation is a method of market segmentation **where the consumer's location of residence is used as the only criterion to form market segments.** The key advantage of geographic segmentation is that it is easy to target communication messages and select communication channels to reach the selected geographic segments. However, living in the same country or area does not necessarily mean that people share other characteristics relevant to marketers, such as their preferred benefits when purchasing a product. Geographic segmentation has been experiencing a revival in international market segmentation studies that aim to extract market segments across geographic boundaries. However, such an approach is challenging because the segmentation variable(s) must be meaningful across all the included geographic regions, and because of known biases that can occur if surveys are completed by respondents from different cultural backgrounds.

▼ Socio-demographic segmentation

Socio-demographic segmentation criteria are based on characteristics such as age, gender, income, and education, which can be used to segment markets. While this approach can be useful in some industries, such as luxury goods, cosmetics, baby products, and retirement villages, where specific socio-demographic characteristics are closely associated with consumer preferences, it has limitations in other areas.

The advantage of socio-demographic segmentation is that segment membership can be easily determined for every consumer, and in some cases, the socio-demographic criterion can explain specific product preferences. For example, having small children is the reason why families choose family vacation villages. However, in many instances, socio-demographic characteristics are not the cause for product preferences and do not provide sufficient market insight for optimal segmentation decisions.

According to Haley (1985), socio-demographic characteristics explain only about 5% of the variance in consumer behavior, which suggests that relying solely on socio-demographic criteria for market segmentation may not provide the best results. Yankelovich and Meer (2006) argue that values, tastes, and preferences are more useful criteria for market segmentation because they are more influential in consumers' buying decisions.

▼ Psychographic Segmentation

Psychographic segmentation is a marketing strategy that groups people based on their psychological traits, beliefs, values, interests, and preferences. It provides a deeper understanding of consumer behavior by identifying the underlying reasons for differences in consumer preferences.

Psychographic segmentation is more complex than other segmentation approaches, such as geographic or demographic, because it requires multiple segmentation variables to be used in order to capture the psychographic dimensions of interest. These segmentation variables could include different travel motives, perceived risks, or other factors that influence consumer behavior.

One advantage of psychographic segmentation is that it provides a more accurate understanding of consumer behavior. For example, if a tourist's primary motivation for going on vacation is to learn about other cultures, they are more likely to undertake a cultural holiday at a destination with cultural treasures to explore. This information can help tourism companies tailor their marketing efforts to target this specific consumer segment.

However, the psychographic approach has its disadvantages. It is more complex and difficult to determine segment memberships for consumers, which can increase the cost and time required to execute effective marketing strategies. Also, the power of the psychographic approach depends heavily on the reliability and validity of the measures used to capture the psychographic dimensions of interest. Therefore, careful attention must be paid to the measurement and validation of the data used in psychographic segmentation studies.

▼ Behavioural Segmentation

- Behavioral segmentation involves grouping people based on their reported or actual behavior
- Advantages of behavioral segmentation include more accurate and meaningful insights
- Brand choice behavior over time has been used as a segmentation variable by several authors
- Using behavioral data avoids the need for developing valid measures for psychological constructs
- Behavioral data may not always be readily available, especially for potential customers
- Behavioral segmentation can be more challenging to implement than other approaches
- Relevant and reliable behavioral data is crucial for meaningful and useful segments
- The context in which the behavior occurs is important to consider
- Behavioral segmentation can be combined with other approaches for a more comprehensive understanding of consumer behavior.

▼ Data from Survey Studies

- Market segmentation analyses are commonly based on survey data.
- Survey data is a cheap and easy way to collect information.
- However, survey data can be biased and lead to inaccurate results.
- There are several types of bias to consider when using survey data, including response bias, social desirability bias, and selection bias.

- To minimize these biases, it's important to design surveys carefully, use appropriate sampling methods, and analyze data thoroughly.

Step4: Exploring Data

▼ A First Glimpse at the Data

- Exploratory data analysis helps to identify measurement levels, investigate univariate distributions, and evaluates dependency structures between variables.
- Data may need to be pre-processed and prepared for use with different segmentation algorithms.
- Results from the data exploration stage provide insights into the suitability of different segmentation methods for extracting market segments.
- A travel motives data set is used to illustrate data exploration in this article.
- The data set contains 20 travel motives reported by 1000 Australian residents in relation to their last vacation.
- The `read.csv()` function is used to read the data set into R, and the resulting data frame is named `vac`.
- The `summary()` function can be used to generate a full summary of the data set or a summary of selected columns.
- The Australian travel motives data set contains answers from 488 women and 512 men, and the age of the participants ranges from 18 to 105 years old.
- The data set contains two income variables, Income and Income2, and Income2 represents a transformation of Income with merged high income categories.

▼ Data Cleaning

- Cleaning data is the first step before data analysis.

- Data cleaning involves checking if all values have been recorded correctly and if consistent labels for the levels of categorical variables have been used.
- **Unlikely values** may indicate errors during data collection or data entry and need to be corrected as part of the data cleaning procedure.
- Levels of categorical variables should contain only permissible values, and any other values need to be corrected.
- Factors are the default format for storing categorical variables in R, and levels of factors are sorted alphabetically.
- R functions like read.csv() or read.table() convert columns containing information other than numbers into factors.
- The possible categories of variables are called levels, and they can be re-ordered.
- Remove duplicates, check for missing data, standardize variables, remove outliers, validate categorical and numeric variables.

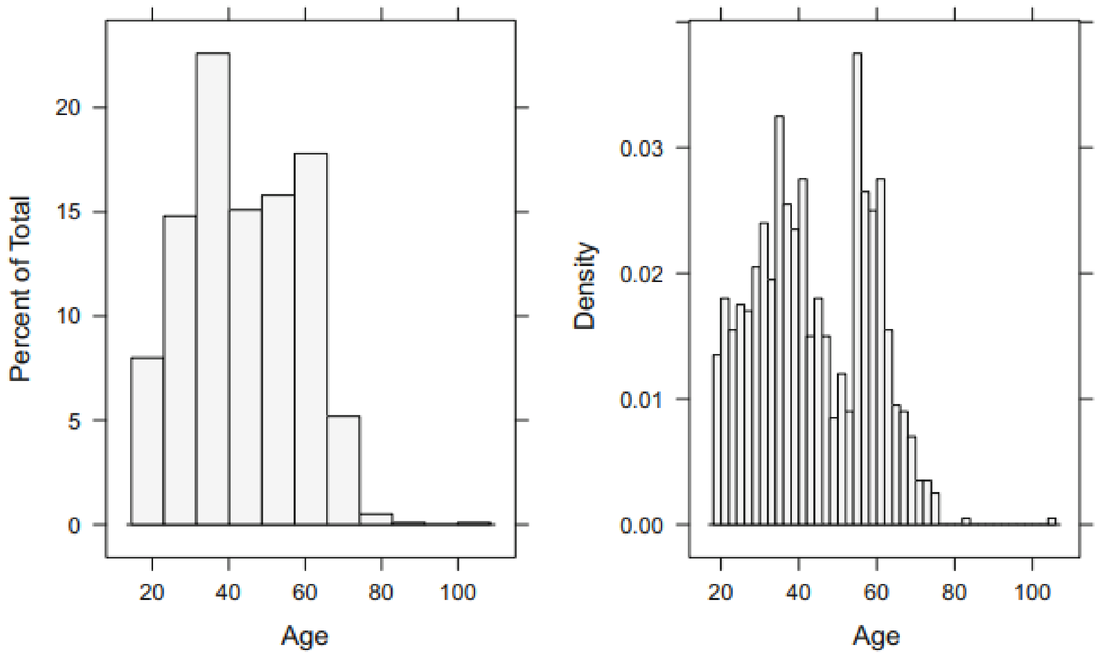
Statistics from the table: The table shows the original and new, re-ordered version of the Income2 variable in the Australian travel motives dataset. The cross-tabulation of the original column with the new, re-ordered version confirms that the transformation was implemented correctly, and the original column of the dataset can safely be overwritten.

▼ Discriptive Statistics

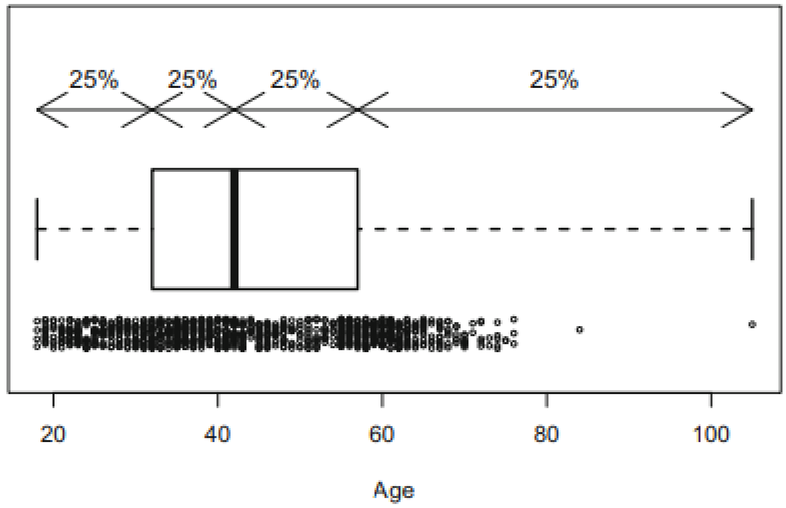
1. Descriptive analysis is important to avoid misinterpretation of results from complex analyses and to gain insights into the data.
2. Statistical software packages, such as R, offer a variety of tools for descriptive analysis including numeric summaries and graphical representations such as histograms, boxplots, and scatter plots for numeric data, and bar plots and mosaic plots for categorical data.
3. Histograms are useful for visualizing the distribution of numeric variables, revealing whether the distribution is unimodal and symmetric or skewed. Binning is necessary to create the histogram, and R packages such as lattice can be used to construct histograms by segments.

Graphical Methods to know the statistics

1. The command "type = 'density'" in the R histogram function rescales the y-axis to display density estimates, allowing for the superimposition of probability density functions of parametric distributions.



2. Box-and-whisker plots are a useful tool for visualizing unimodal distributions of data. Outliers can affect the representation of the data in a boxplot, so most statistical packages impose a restriction on the length of the whiskers to ensure that information about outliers is not lost.



- Preprocessing
  - Categorical Variables

    In data pre-processing for categorical variables, two common procedures are merging levels of categories and converting them to numeric variables. Merging is useful when there are too many categories that are too differentiated. Ordinal data can be converted to numeric data if distances between adjacent scale points are assumed to be equal. Binary answer options are less prone to capturing response styles, and do not require pre-processing. Binary variables can always be converted to numeric variables, and most statistical procedures work correctly after conversion if there are only two categories.

  - Numerical Variables

    In distance-based segmentation methods, the range of values of a segmentation variable can affect its influence on the segmentation results. Standardizing variables can balance their influence by transforming them to a common scale. The default method subtracts the mean and divides by the standard deviation. Standardization can be done in R using the scale() function. Alternative methods may be needed for data with outliers. In such cases, robust estimates like the median and interquartile range are preferred.

Z-score formula:

$z = (x - \mu) / \sigma$

standard deviation formula:

$s = \text{sqrt}((1/(n-1)) * \Sigma(xi - \bar{x})\text{^}2)$

▼ Principal Components Analysis

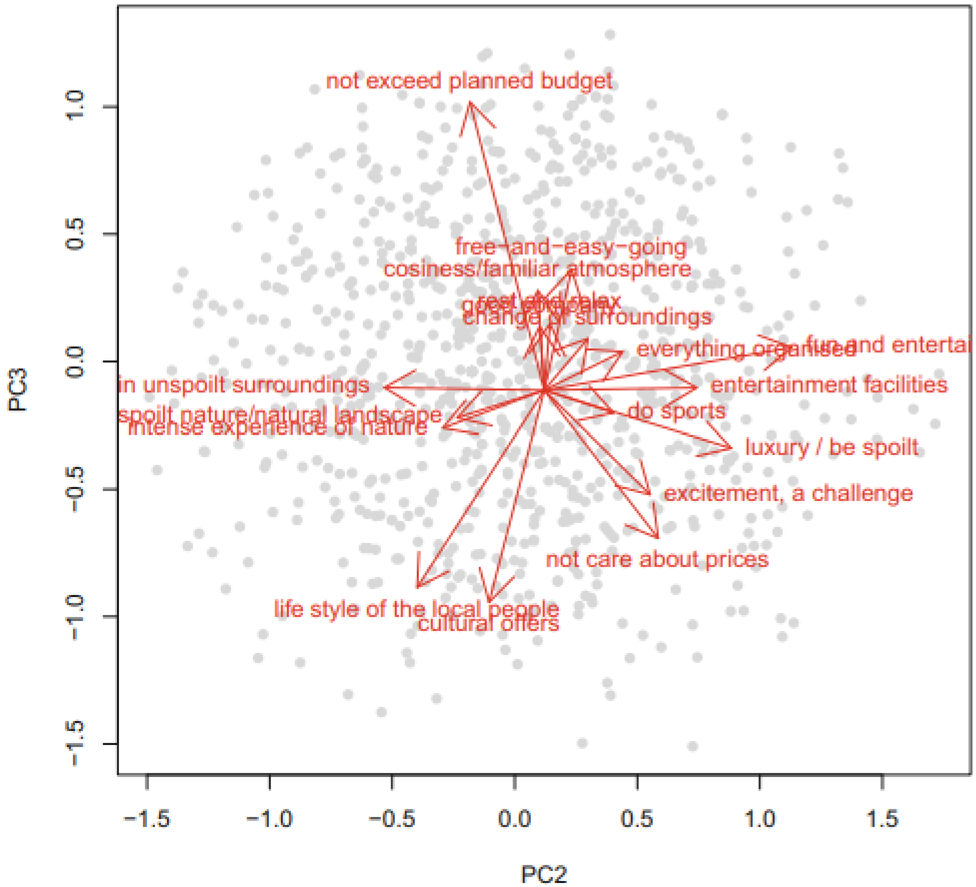
Principal Component Analysis (PCA) is a statistical method used to transform a multivariate data set containing metric variables to a new data set with variables called principal components. These principal components are uncorrelated and are ordered by their importance. The first principal component contains most of the variability, the second principle component contains the second most variability, and so on.

PCA works on the covariance or correlation matrix of several numeric variables. If all variables are measured on the same scale and have similar data ranges, it is not important which one to use. If the data ranges are different, the correlation matrix should be used (which is equivalent to standardising the data).

The transformation obtained from PCA is commonly used to project high-dimensional data into lower dimensions for plotting purposes. In this case, only a subset of principal components is used, typically the first few because they capture the most variation. The first two principal components can easily be inspected in a scatter plot, and more than two principal components can be visualised in a scatter plot matrix.

In R, we can use the `prcomp()` function to perform PCA on a dataset. For example, the following command generates a principal components analysis for the Australian travel motives data set:

```
vacmot.pca <- prcomp(vacmot)
```



By default, `prcomp()` centers the data but does not standardize it. We can inspect the resulting object `vacmot.pca` by printing it. The print output shows the standard deviations of the principal components, which reflect the importance of each principal component. The print output also shows the rotation matrix, which specifies how to rotate the original data matrix to obtain the principal components.

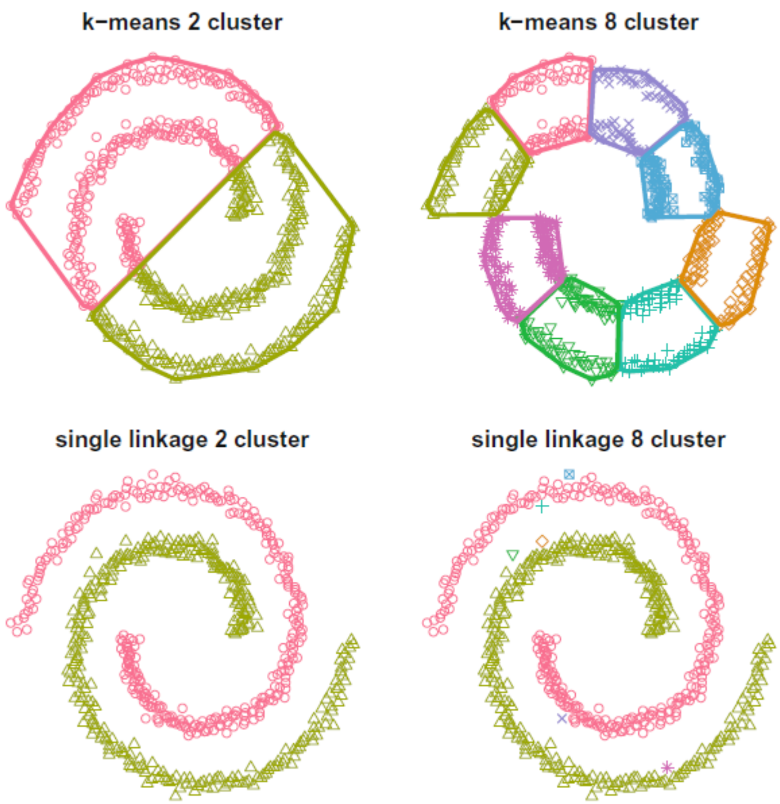
We can obtain further information on the fitted object with the `summary()` function. For objects returned by function `prcomp()`, the function `summary()` provides standard deviation, proportion of variance and cumulative proportion of variance for each principal component.

Step5: Extracting Segments

▼ Grouping consumers

Data-driven market segmentation analysis is exploratory by nature. Consumer data sets are typically not well structured. The combination of exploratory methods and unstructured consumer data means that results from any method used to extract market segments from such data will strongly depend on the assumptions made on the structure of the segments implied by the method. Therefore, the result of a market segmentation analysis is determined as much by the underlying data as it is by the extraction algorithm chosen. Segmentation methods shape the segmentation solution.

Many segmentation methods used to extract market segments are taken from the field of cluster analysis. It is, therefore, important to explore market segmentation solutions derived from a range of different clustering methods. Understanding how different algorithms impose structure on the extracted segments is also important.



In the above diagram, the top row shows the market segments obtained when running k-means cluster analysis, and the bottom row shows the market segments obtained from single linkage hierarchical clustering.

There is no single best algorithm for all data sets. If consumer data is well-structured and well-separated, distinct market segments exist, and the 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 algorithm's objective function.

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

The above table contains the information needed to guide algorithm selection. Data characteristics and expected or desired segment characteristics allow a pre-selection of suitable algorithms to be included in the comparison.

In the case of binary segmentation variables, another aspect needs to be considered. We may want consumers in the same segments to have both the presence
and absence of segmentation variables in common. In this case, we need to treat the binary segmentation variables symmetrically (with 0s and 1s treated equally). Alternatively, we may only care about segmentation variables consumers have in common. In this case, we treat them asymmetrically (with only common 1s being of interest).

▼ Distance-Based methods

Market segmentation aims at grouping consumers into groups with similar needs or behavior, in this example: groups of tourists with similar patterns of vacation activities.

In order to find groups of similar tourists one needs a notion of similarity or dissimilarity, mathematically speaking: a distance measure.

▼ Distance Measures

This is a typical data matrix. Each row represents an observation and every column represents a variable.

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

The vector corresponding to the i-th row of matrix X is denoted as xi = (xi1, xi2, . . . , xip)’ in the following, such that X = {x1, x2, . . . xp} is the set of all observations.

The most common distance measures used in market segmentation analysis are:
Euclidean distance, Manhattan or absolute distance, Asymmetric binary distance.

▼ Hierarchical Methods

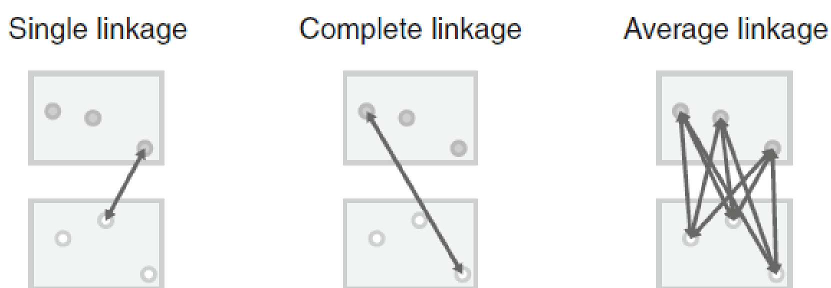
Hierarchical clustering methods are the most intuitive way of grouping data because they mimic how a human would approach the task of dividing a set of observations n (consumers) into groups (segments). If the aim is to have one large market segment (k = 1), the only possible solution is one big market segment containing all consumers in the data X. At the other extreme, if the

aim is to have as many market segments as there are consumers in the data set (k = n), the number of market segments has to be n, with each segment containing exactly one consumer. Each consumer represents their own cluster. Market segmentation analysis occurs between those two extremes.
Divisive hierarchical clustering methods start with the complete data set X and splits it into two market segments in the first step. Then, each of the segments is again split into two segments. This process continues until each consumer has their own market segment.
Agglomerative hierarchical clustering approaches the task from the other end. The starting point is each consumer representing their own market segment

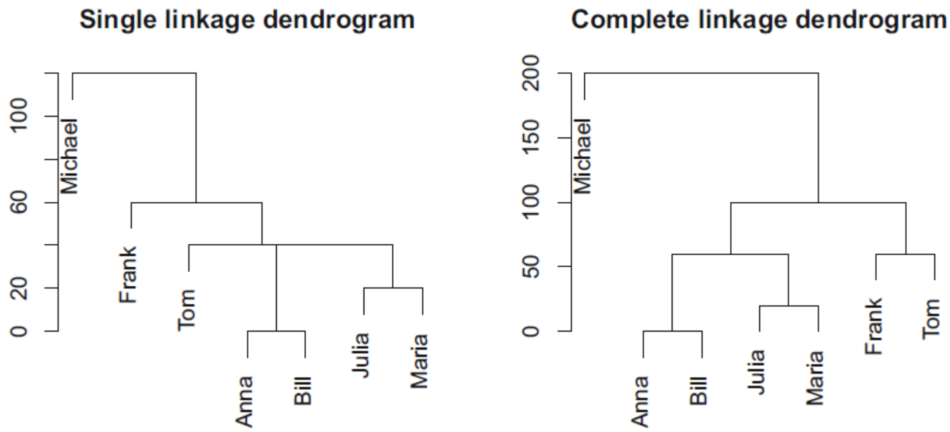
Single linkage: Distance between the two closest observations of the two sets

Complete linkage: Distance between the two observations of the two sets that are farthest away from each other

Average linkage: Mean distance between observations of the two sets

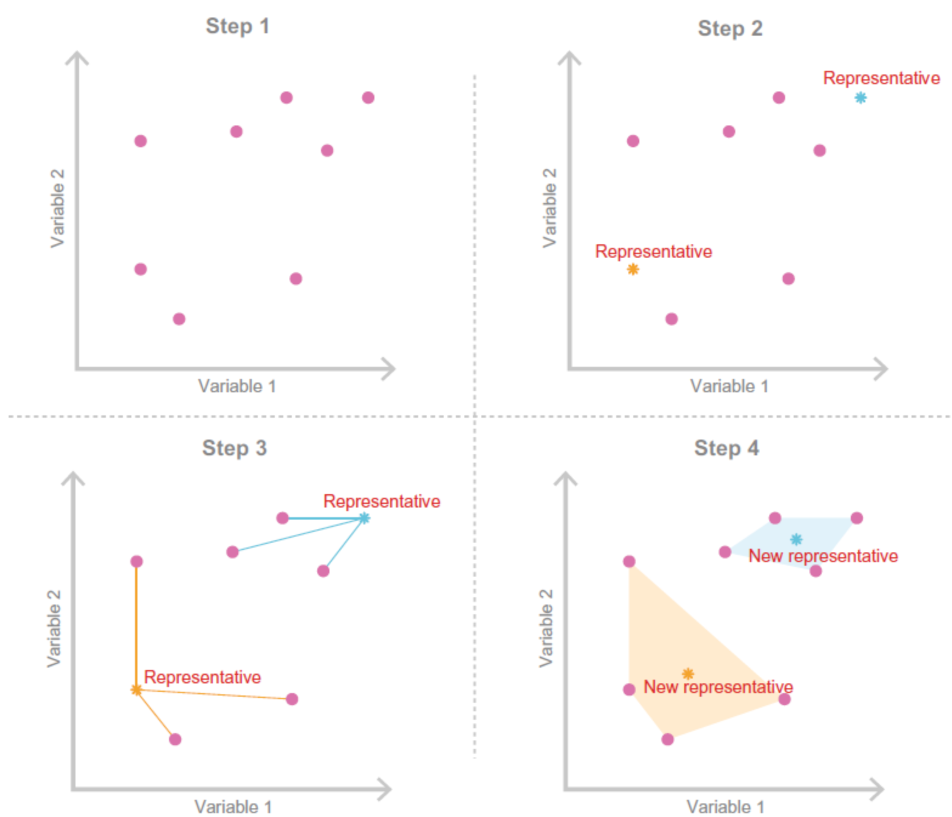


The result of hierarchical clustering is typically presented as a dendrogram. A dendrogram is a tree diagram. The root of the tree represents the one-cluster solution where one market segment contains all consumers. The leaves of the tree are the single observations (consumers), and the branches in between 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 selecting the number of market segments.



▼ Partitioning Methods

1. k-means and k-centroid clustering



Simplified visualization of the k-means clustering algorithm

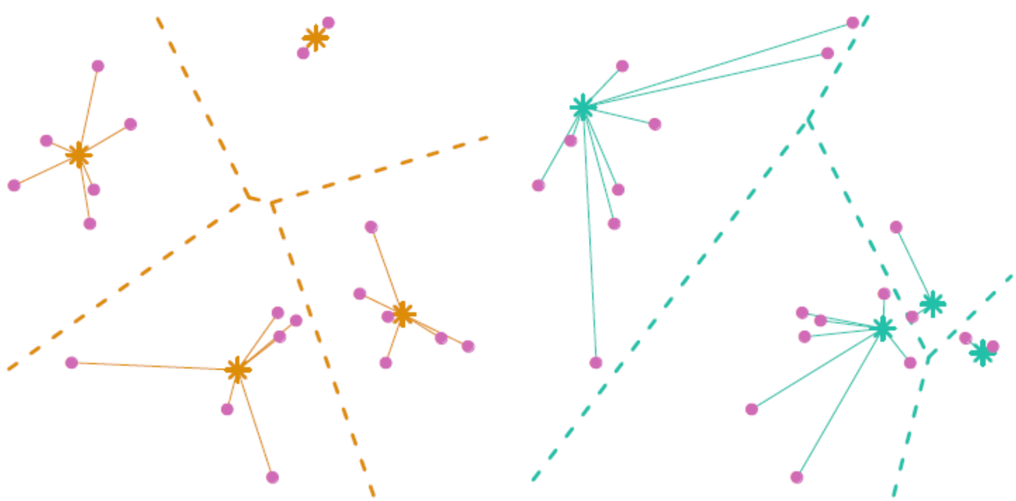
The first thing that this algorithm does is it will take a random guess at where might be the centers of the two clusters that you might ask to find.

It will pick random two points that where might be the centers of two different clusters.

2. Improved k-means

   Many attempts have been made to refine and improve the k-means clustering
   algorithm. The simplest improvement is to initialize k-means using “smart” starting values, rather than randomly drawing k consumers from the data set and using them as starting points. Using randomly drawn consumers is suboptimal because it may result in some of those randomly drawn consumers being located very close to one another, and thus not being representative of the data space. Using starting points that are not representative of the data space increases the likelihood of the k-means algorithm getting stuck in what is referred to as a local optimum. A local optimum is a good solution, but not the best possible solution. One way of avoiding the problem of the algorithm getting stuck in a local optimum is to initialize it using starting points evenly spread across the entire data space. Such starting points better represent the entire data set.

3. Hard Competitive Learning



Examples of good (left) and bad (right) starting points for k-means clustering

each consumer contained in the data set to their closest representative (centroid), the process by which this is achieved is slightly different. k-means uses all consumers in the data set at each iteration of the analysis to determine the new segment representatives (centroids). Hard competitive learning randomly picks one consumer and moves this consumer’s closest segment representative a small step in the direction of the randomly chosen consumer.

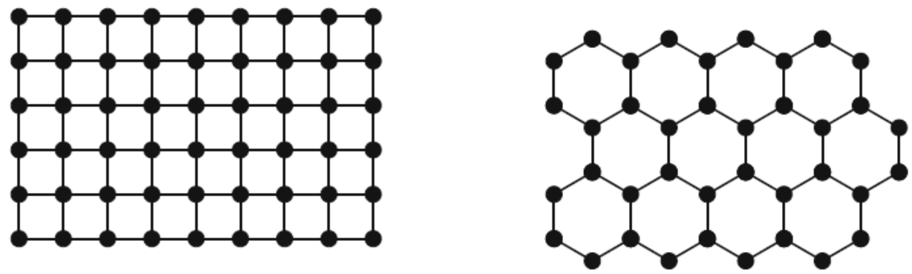
4. Neural Gas and Topology Representing Networks

   A variation of hard competitive learning is the neural gas algorithm proposed by Martinetz et al. (1993). Here, not only the segment representative (centroid) is moved towards the randomly selected consumer. Instead, also the location of the second closest segment representative (centroid) is adjusted towards the randomly selected consumer. However, the location of the second closest representative is adjusted to a smaller degree than that of the primary representative.

5. Self-organizing maps

   The self-organizing map algorithm is similar to hard competitive learning: a single random consumer is selected from the data set, and the closest representative for this random consumer moves a small step in their direction.

   In addition, representatives which are direct grid neighbors of the closest representative move in the direction of the selected random consumer. The process is repeated many times; each consumer in the data set is randomly chosen multiple times and used to adjust the location of the centroids in the Kohonen map.



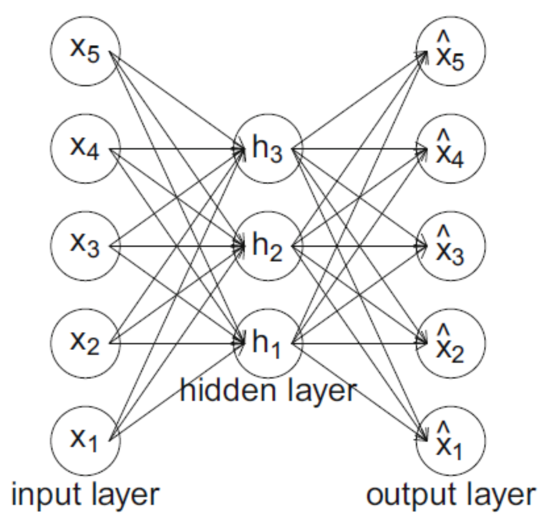
Rectangular (left) and hexagonal (right) grid for self-organizing maps

6. Neural Networks

   Auto-encoding neural networks for cluster analysis work mathematically differently than all cluster methods presented so far. The most popular method from this family of algorithms uses a so-called single hidden layer perceptron.

   The values of the three nodes in the hidden layers h1, h2, and h3 are weighed in linear combinations of the inputs.

$$h_j = f_j\left(\sum_{i=1}^{5} \alpha_{ij} x_i\right)$$



▼ Hybrid approaches

Several approaches combine hierarchical and partitioning algorithms in an attempt
to compensate for the weaknesses of one method with the strengths of the other.
The strengths of hierarchical cluster algorithms are that the number of market segments to be extracted does not have to be specified in advance and that similarities of market segments can be visualized using a dendrogram. The biggest
disadvantage of hierarchical clustering algorithms is that standard implementations
require substantial memory capacity, thus restricting the possible sample size of the data for applying these methods. Also, dendrograms become very difficult to interpret when the sample size is large.

1. Two-step clustering

   The two steps consist of running a partitioning procedure followed by a hierarchical procedure.

2. Bagged Clustering

   combines hierarchical clustering algorithms and partitioning clustering algorithms, but adds bootstrapping (Efron and Tibshirani 1993). Bootstrapping can be implemented by random drawing from the data set with replacement. That means that the process of extracting segments is repeated many times with randomly drawn (bootstrapped) samples of the data. Bootstrapping has the advantage of making the final segmentation solution less dependent on the exact people contained in consumer data.

   Bagged clustering is suitable for the following circumstances:

   - If we suspect the existence of niche markets.
   - If we fear that standard algorithms might get stuck in bad local solutions.
   - If we prefer hierarchical clustering, but the data set is too large

▼ Model-based methods

Here, a slightly more pragmatic perspective is taken. Model-based methods are viewed as one additional segment extraction method available to data analysts. Given that extracting market segments is an exploratory exercise, it is helpful to use a range of extraction methods to determine the most suitable approach for the data at hand. Having model-based methods available is particularly useful because these
methods extract market segments in a very different way, thus genuinely offering an alternative extraction technique.

▼ Finite Mixtures of Distributions

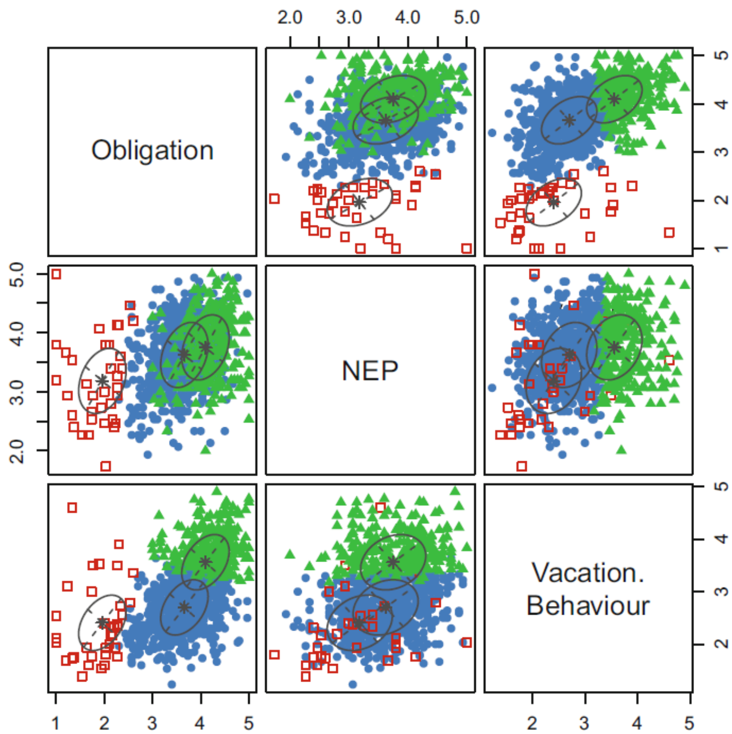
1. Normal Distributions

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

   Mathematically, f () in above fig is the multivariate normal distribution which has
   two sets of parameters (mean and variance) like the univariate normal distribution. If p segmentation variables are used, these have p mean values, and each segment has a segment-specific mean vector μh of length p. In addition to the p variances of the p segmentation variables, the covariance structure can be modeled, resulting in a p×p covariance matrix summation of h for each segment. The covariance matrix summation of h contains the variances of the p segmentation variables in the diagonal and the covariances between pairs of segmentation variables in the other entries. The covariance matrix is symmetric and contains p(p + 1)/2 unique values.
   The segment-specific parameters θh are the combination of the mean vector μh
   and the covariance matrix summation of h, and the number of parameters to estimate is p + p(p + 1)/2.

2. Binary Distributions

   The mixture model assumes that respondents in different segments have different probabilities of undertaking certain activities. For example, some respondents may be interested in alpine skiing and not interested in sightseeing. This leads to these two variables being negatively correlated in the overall data set. However, this correlation is due to groups of respondents interested in one of the two activities only.



Classification plot of the mixture of normal distributions for the Australian travel motives data set selected using the BIC among the models with identical covariance matrices across segments

- ▼ Finite Mixtures of Regressions

  Finite mixtures of distributions are similar to distance-based clustering methods and – in many cases – result in similar solutions. Compared to hierarchical or partitioning clustering methods, mixture models sometimes produce more useful,
  and sometimes less useful solutions.

- ▼ Extensions and Variations

▼ Algorithms with integrated Variable Selection

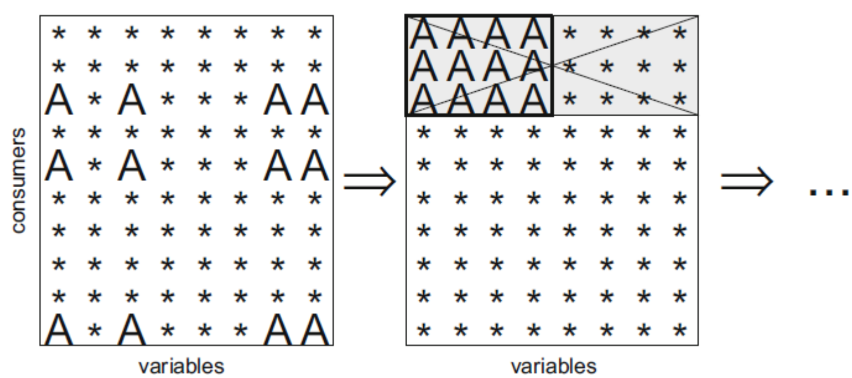
- ▼ Bi-clustering Algorithms

Biclustering simultaneously clusters both consumers and variables. Biclustering algorithms exist for any kind of data, including metric and binary. This section focuses on the binary case where these algorithms aim at extracting market segments containing consumers who all have a value of 1 for a group of variables. These groups of consumers and variables together then form the bicluster.

The biclustering algorithm which extracts these biclusters follows a sequence of steps. The starting point is a data matrix where each row represents one consumer
and each column represents a binary segmentation variable:

Step1:

First, rearrange rows (consumers) and columns (segmentation variables) of the data matrix in a way to create a rectangle with identical entries of 1s at the top left of the data matrix. The aim is for this rectangle to be as large as possible.



Biclustering with constant pattern

Step 2:

Second, assign the observations (consumers) falling into this rectangle to one bicluster, as illustrated by the grey shading in Fig. The segmentation variables defining the rectangle are active variables (A) for this bicluster.

Step 3:

Remove from the data matrix the rows containing the consumers who have been assigned to the first bicluster. Once removed, repeat the procedure from step 1 until no more biclusters of sufficient size can be located.

▼ Variable Selection Procedure for Clustering Binary Data

VSBD method is based on the k-means algorithm as clustering method, and assumes that not all variables available are relevant to obtain a good clustering solution. In particular, the method assumes the presence of masking variables. They need to be identified and removed from the set of segmentation variables. Removing irrelevant variables helps to identify the correct segment structure, and
eases interpretation.

The algorithm works as follows:

Step 1:

Select only a subset of observations with size $\varphi \in (0, 1]$ times the size of the original data set. Brusco (2004) suggests to use $\varphi = 1$ if the original data set contains less than 500 observations, $0.2 \leq \varphi \leq 0.3$ if the number of observations is between 500 and 2000 and $\varphi = 0.1$ 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 within-cluster 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-of-squares value if added to the set of segmentation variables.

Step 4:

Add this variable if the increase in within-cluster sum-of-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 Reduction: Factor-Cluster Analysis

  The term factor-cluster analysis refers to a two-step procedure of data-driven market
  segmentation analysis. In the first step, segmentation variables are factor analysed.
  The raw data, the original segmentation variables, are then discarded. In the second
  step, the factor scores resulting from the factor analysis are used to extract market
  segments.

▼ Data-Structure Analysis

- ▼ Cluster Indices

  Because market segmentation analysis is exploratory, data analysts need guidance to make some of the most critical decisions, such as selecting the number of market segments to extract. So-called cluster indices represent the most common approach to obtaining such guidance. Cluster indices provide insight into particular aspects of the market segmentation solution. Which kind of insight, depends on the nature of the clustered index used. Generally, two groups of cluster indices are distinguished: internal cluster indices and external cluster indices.

  1. Internal Cluster Indices

     Internal cluster indices ask one of two questions or consider their combination: (1) how compact is each of the market segments? and (2) how well separated are different market segments? To answer these questions, the notion of a distance measure between observations or groups of observations is required.

  2. External Cluster Indices

External cluster indices evaluate a market segmentation solution using additional
external information; they cannot be calculated using only the information contained in one market segmentation solution. A range of different additional pieces of information can be used. The true segment structure – if known – is the most valuable additional piece of information. But the true segment structure of the data is typically only known for artificially generated data. The true segment
structure of consumer data is never known. When working with consumer data,
the market segmentation solution obtained using a repeated calculation can be used as additional, external information. The repeated calculation could use a different clustering algorithm on the same data, or it could apply the same algorithm to a variation of the original data

▼ Gorge Plots

A simple method to assess how well segments are separated is to look at the distances of each consumer to all segment representatives. Let dih be the distance
between consumer i and segment representative (centroid, cluster centre) h. They can be interpreted as the similarity of consumer i to the representative of segment h, with hyperparameter γ controlling how differences in distance translate into differences in similarity. These similarities are between 0 and 1, and sum to 1 for each consumer i over all segment representatives h, h = 1, . . . , k.

$$s_{ih} = \frac{e^{-d_{ih}^{\gamma}}}{\sum_{l=1}^{k} e^{-d_{il}^{\gamma}}}$$

For partitioning methods, segment representatives and distances between consumers and segment representatives are directly available. For model-based methods, we use the probability of a consumer i being in segment h given the consumer data, and the fitted mixture model to assess similarities. In the mixture of normal distributions case, these probabilities are close to the similarities obtained with Euclidean distance and γ = 2 for k-means clustering. Below we use γ = 1 because it shows more details, and led to better results in simulations on artificial data. The parameter can be specified by the user in the R implementation.

- ▼ Global Stability Analysis

  An alternative approach to data structure analysis that can be used for both distance and model-based segment extraction techniques is based on resampling methods. Resampling methods offer insight into the stability of a market segmentation solution across repeated calculations. To assess the global stability of any given segmentation solution, several new data sets are generated using resampling methods, and a number of segmentation solutions are extracted.

- ▼ Segment Level Stability Analysis

  Choosing the globally best segmentation solution does not necessarily mean that
  this particular segmentation solution contains the single best market segment. Relying on global stability analysis could lead to selecting a segmentation solution
  with suitable global stability, but without a single highly stable segment. It is recommendable, therefore, to assess not only the global stability of alternative market segmentation solutions but also the segment level of stability of market segments contained in those solutions to protect against discarding solutions containing interesting individual segments from being prematurely discarded. After all, most organizations only need one single target segment.

  - ▼ Segment-level Stability within solutions

    The criterion of segment-level stability within solutions (SLSW) is similar to the concept of global stability. The difference is that stability is computed at the segment level, allowing the detection of one highly stable segment (for example a potentially attractive niche market) in a segmentation solution where
    several or even all other segments are unstable.

    Hennig recommends the following steps:

1. Compute a partition of the data (a market segmentation solution) extracting k segments S1, . . . ,Sk using the algorithm of choice (for example, a partitioning clustering algorithm or a finite mixture model).
2. Draw b bootstrap samples from the sample of consumers including as many cases as there are consumers in the original data set (b = 100 bootstrap samples works well).
3. Cluster all b bootstrap samples into k segments. Based on these segmentation solutions, assign the observations in the original data set to segments Si 1, . . . ,Sik for i = 1, . . . , b.
4. For each bootstrap segment Si1, . . . ,Sik, compute the maximum agreement with the original segments S1, . . . ,Sk as measured by the Jaccard index:

$$s_h^i = \max_{1 \leq h' \leq k} \frac{|\mathcal{S}_h \cap \mathcal{S}_{h'}^i|}{|\mathcal{S}_h \cup \mathcal{S}_{h'}^i|}, \qquad 1 \leq h \leq k.$$

   The Jaccard index is the ratio between the number of observations contained in both segments, and the number of observations contained in at least one of the two segments.
5. Create and inspect boxplots of the sih values across bootstrap samples to assess the segment level stability within solutions (SLSW). Segments with higher segment level stability within solutions (SLSW) are more attractive.

▼ Segment-level stability across solutions

The second criterion of stability at segment level proposed by Dolnicar and Leisch (2017) is referred to as segment level stability across solutions (SLSA). The purpose of this criterion is to determine the re-occurrence of a market segment across market segmentation solutions containing different numbers of segments. High values of segment level stability across solutions (SLSA) serve as indicators of market segments occurring naturally in the data, rather than being artificially created. Natural segments are more attractive to organisations because they actually exist, and no managerial judgement is needed in the artificial construction of segments.

Let P1, . . . ,Pm be a series of m partitions (market segmentation solutions) with
kmin, kmin +1, kmin +2, . . . , kmax segments, where m = kmax −kmin +1. The
minimum and maximum number of segments of interest (kmin and kmax) have to be specified by the user of the market segmentation analysis in collaboration with the data analyst.

Step6: Profiling Segments

```
import flexclust
from flexclust import data
data(vacmot, package="flexclust")
with open("vacmot-clusters.RData", "rb") as f:
    vacmot_clusters = pickle.load(f)



from scipy.spatial.distance import pdist
from scipy.cluster.hierarchy import ward, dendrogram

vacmot_vdist = pdist(vacmot.T)
vacmot_vclust = ward(vacmot_vdist)



import matplotlib.pyplot as plt

# assuming vacmot.k6 is a vector of cluster assignments
# and vacmot.vclust is a hierarchical clustering object
# with the order of the leaves as vacmot.vclust.order
order = vacmot_vclust.order[::-1]
barchart_data = vacmot_k6
plt.bar(range(len(barchart_data)), barchart_data[order], color='gray')
plt.xticks(range(len(barchart_data)), order)
plt.show()
```

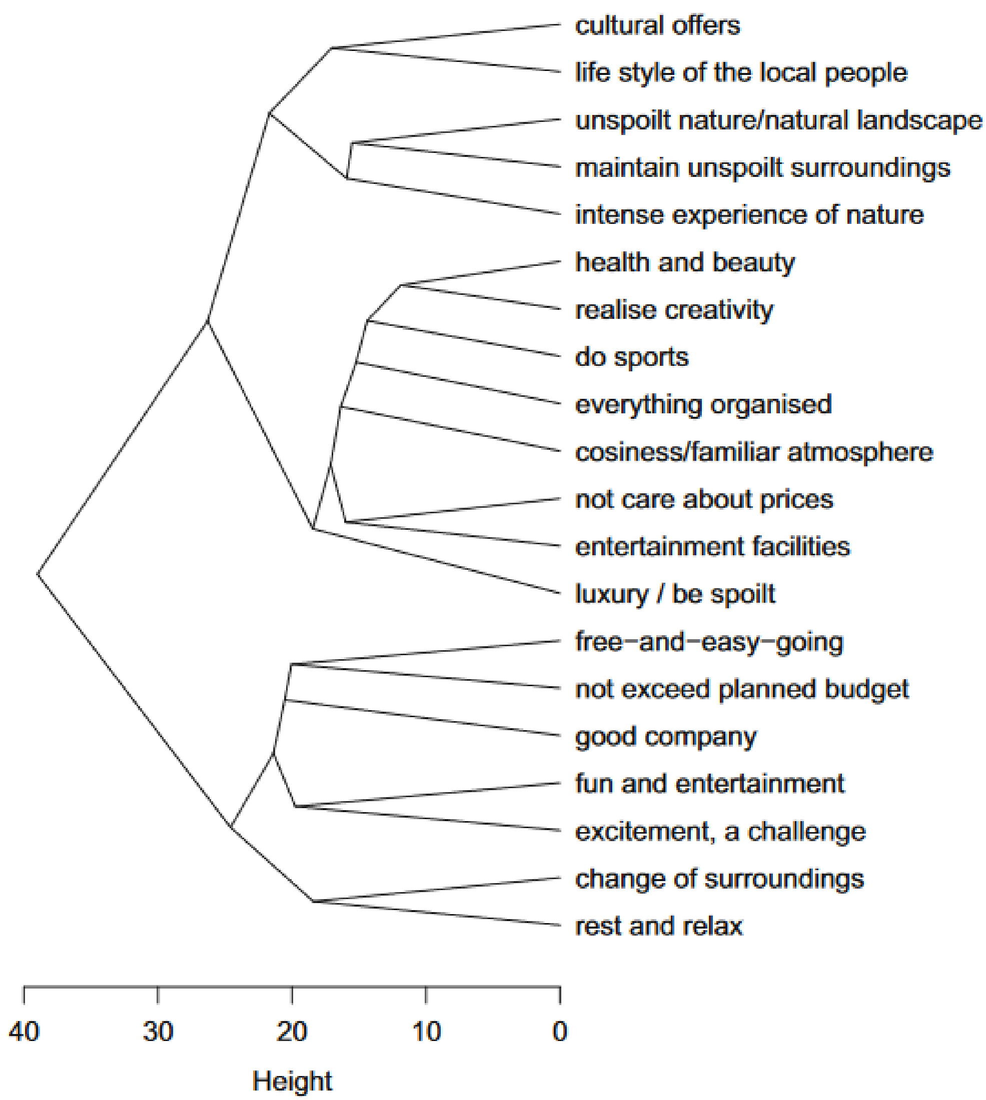
Identifying Key Characteristics

- For commonsense segmentation, the pro□les of the segments are predefined, eg. age

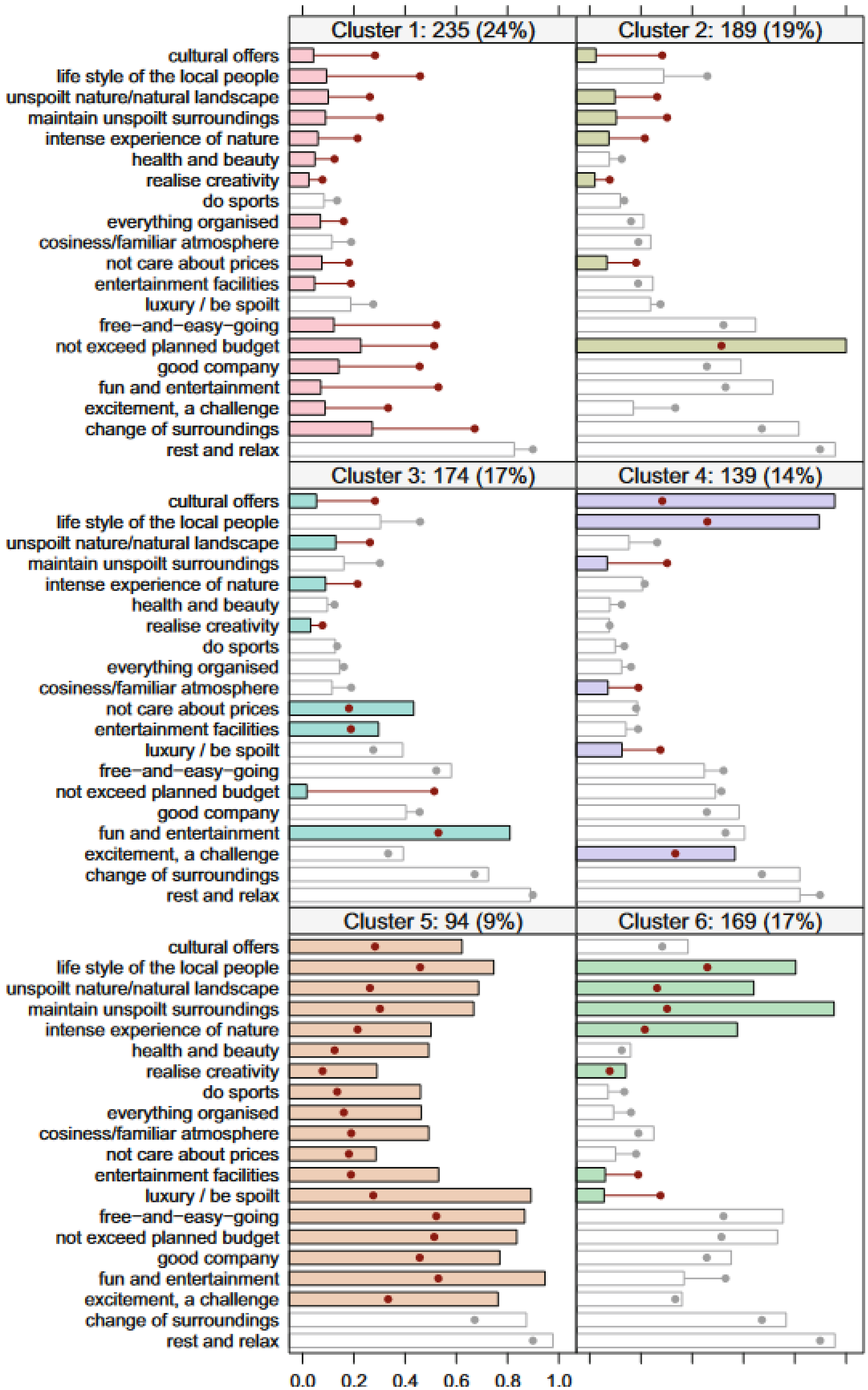
- Profling consists of characterising the market segments individually, but also in comparison to the other market segments
- we inspect a number of alternative market segmentation solutions. This is particularly important if no natural segments exist in the data
- Data-driven segmentation solutions are usually presented to users (clients, managers) in one of two ways:
  - as high level summaries simplifying segment characteristics to a point where they are misleadingly trivial, or
  - 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.

Segment Profiling with Visualizations

- Graphics are particularly important in exploratory statistical analysis (like cluster analysis)
- 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.
- Another option is to order segmentation variables by similarity of answer patterns. We can achieve this by clustering the columns of the data matrix
- hierarchical clustering of the variables is conducted using Ward’s method


cultural offers
life style of the local people
unspoilt nature/natural landscape
maintain unspoilt surroundings
intense experience of nature
health and beauty
realise creativity
do sports
everything organised
cosiness/familiar atmosphere
not care about prices
entertainment facilities
luxury / be spoilt
free−and−easy−going
not exceed planned budget
good company
fun and entertainment
excitement, a challenge
change of surroundings
rest and relax
40
30
20
10
0
Height

•



Cluster 1: 235 (24%)
Cluster 2: 189 (19%)
Cluster 3: 174 (17%)
Cluster 4: 139 (14%)
Cluster 5: 94 (9%)
Cluster 6: 169 (17%)
cultural offers
life style of the local people
unspoilt nature/natural landscape
maintain unspoilt surroundings
intense experience of nature
health and beauty
realise creativity
do sports
everything organised
cosiness/familiar atmosphere
not care about prices
entertainment facilities
luxury / be spoilt
free−and−easy−going
not exceed planned budget
good company
fun and entertainment
excitement, a challenge
change of surroundings
rest and relax
0.0
0.2
0.4
0.6
0.8
1.0

Step 7: Describing Segments

- Developing a Complete Picture of Market Segments

  The marketing literature traditionally relies on statistical testing, and tabular presentations of differences in descriptor variables. Visualisations make segment description more user-friendly.

- Using Visualizations to Describe Market Segments

  Graphical statistics to describe market segments has two key advantages: simplifying interpretation and integrating statistical significance, and conveying the essence of marketing research results.

  Metric Descriptor Variables

  Conditional plots are useful for visualising differences between market segments using metric descriptor variables, such as age distribution, moral obligation scores, and segment names.

  The most important idea is that members of the acquiescence segment have an overall tendency to express agreement with survey questions, and are more likely to express agreement when asked about their moral obligation to protect the environment.

- Testing for Segment Differences in Descriptor Variables

  The $\chi 2$-test rejects the null hypothesis of independence if the p-value is smaller than 0.05, so a mosaic plot is used to identify the reason for rejection. ANOVA is used to test for significant differences in mean moral obligation values across market segments.
  The analysis of variance performs an F-test to compare the weighted variance between market segment means with the variance within market segments. Small values support the null hypothesis that segment means are the same, but the analysis of variance does not identify the differing segments. Pairwise comparisons between segments provide this information.
  P-values need to be adjusted for multiple testing when assessing a single hypothesis, such as that all segment means are the same. Bonferroni correction is

the simplest way to correct p-values for multiple testing, as it multiplies all p-values by the number of tests computed.

- Predicting Segments from Descriptor Variables

Regression analysis is the basis of prediction models, which test differences in all descriptor variables simultaneously to determine how well members of a market segment can be identified.
Segment membership C6 is a categorical variable with six categories, and the formula interface fits a regression coefficient for each category. Without an intercept, each estimated coefficient is equal to the mean age in the segment.

Binary Logistic Regression

$$g(\mu) = \eta = \log\left(\frac{\mu}{1-\mu}\right)$$

The binomial distribution is a generalisation of the Bernoulli distribution if the variable *y* does not only take values 0 and 1, but represents the number of successes out of a number of independent Bernoulli distributed trials with the same success probability *μ*.

The regression coefficients in a linear regression model indicate how much the mean value of the dependent variable changes if this independent variable changes while others remain unchanged. In binary logistic regression, the regression coefficients indicate how the linear predictor changes. The coefficient for AGE indicates that the log odds for being in segment 3 are 0.008 lower for tourists who are one year older. The independent variable OBLIGATION2 is a categorical variable with four different levels. The regression coefficients for this variable indicate the change in log odds between the other categories and the lowest category Q1.

To simplify the interpretation of the coefficients and their effects, we can use package effects (Fox 2003; Fox and Hong 2009) in R. Function allEffects calculates the predicted values for different levels of the independent variable keeping other independent variables constant at their average value. The predicted values are the probabilities of being in the segment 3. We plot the estimated probabilities to allow for easy inspection.

Multinomial Logistic Regression

In R, function multinom() from package fits a multinomial logistic regression. The regression coefficients are arranged in matrix form. Each row contains the regression coefficients for one category of the dependent variable. Each column contains the regression coefficients for one effect of an independent variable.

Tree-Based Methods

Classification and regression trees (CARTs) are an alternative modelling approach for predicting a binary or categorical dependent variable given a set of independent variables. They use a stepwise procedure to fit the model, splitting consumers into groups based on one independent variable. The resulting tree shows the nodes that emerge from each splitting step, with the root node containing all consumers and terminal nodes that are not split further. Segment membership can be predicted based on the segment memberships of consumers contained in the terminal node. Package partykit implements unbiased variable selection and enables visualisation of fitted tree models.

The output of the fitted classification tree shows that consumers with a Vacation. Behaviour score of 2.2 or less are assigned to node 2, while consumers with a score higher than 2.2 are assigned to Node 3. Node 4 contains 490 respondents, 81% of whom are not in segment 3, 19% are, and most of them are in node 5. Plotting the classification tree using plot(tree63) gives a visual representation that is easier to interpret. Consumers with a value higher than 2.2 follow the right branch to node 3, while those with a value of 2.2 or less follow the left branch to node 2. The proportion of respondents in node 2 who belong to segment 3 is shown at the bottom of the stacked bar chart.

The output shows that the first splitting variable is the categorical variable indicating moral obligation (OBLIGATION2). This variable splits the root node 1 into nodes 2 and 5. Node 2 is split into nodes 3 and 4 using EDUCATION as splitting variable. Node 3 is a terminal node and contains 481 respondents. Node 4 contains 286 respondents and 77% of them are not in segment 1. Node 5 contains respondents with a moral obligation value of 47 or less, and a moral obligation category value of Q4. Node 6 contains 203 respondents and 67% are not from segment 6. Node 7 contains 30 consumers and 57% do not belong to segment 5. Most of the plot is the same as for the classification tree with the binary dependent variable, except for the bar charts at the bottom. Optimally, these bar charts for each terminal node show that nearly all consumers in that node have the same segment membership or are at least assigned to only a small number of different segments.

Step8: Selecting the Target Segment

- The Targeting Decision

  Market targeting is a critical decision that significantly affects an organization's future performance. It involves selecting one or more market segments that an organization wants to focus on and committing to them for the long term.

  Before selecting a target segment, an organization must first identify and agree upon knock-out criteria for market segments. These criteria eliminate any segments that are too small, not homogeneous, not distinct enough, or have needs that the organization cannot meet.

  Once the knock-out criteria have been applied, the remaining market segments' attractiveness and the organization's competitiveness for each segment must be evaluated. The organization must determine which segment it most wants to target and commit to, as well as which organization offering the same product each segment would most likely choose.

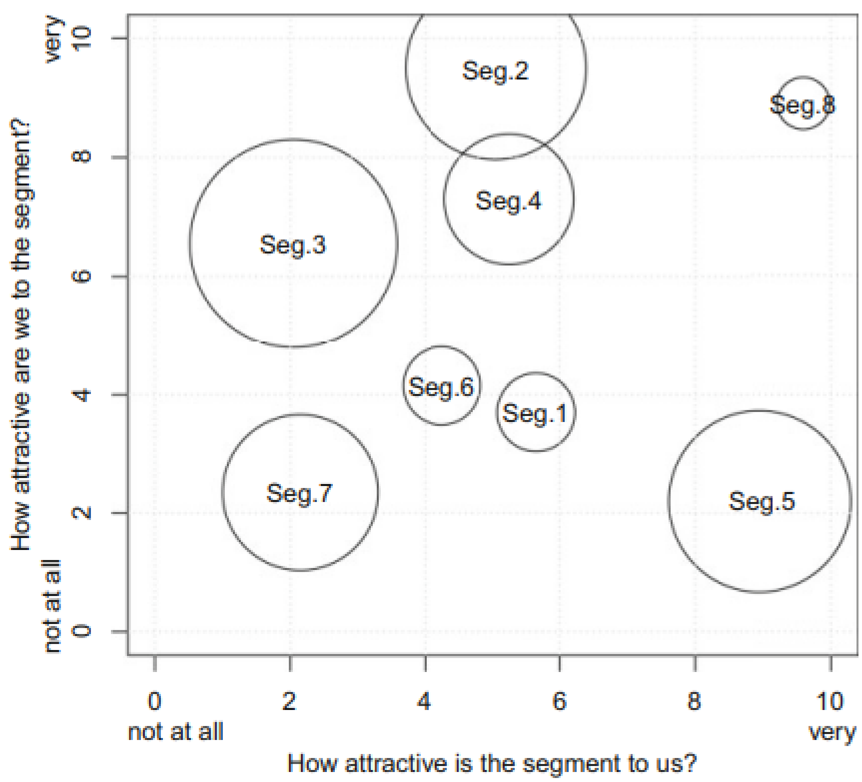
  Answering these two questions forms the basis of the target segment decision, which is a critical step in strategic marketing. The organization's success in this decision will significantly affect its future performance.

- Market Segment Evaluation

  The selection of a target market is an essential component of marketing strategy. To evaluate different market segments, a decision matrix is often recommended by many experts. The purpose of this matrix is to visualize the relative attractiveness of the segment and organizational competitiveness to assist with decision-making. This report will discuss various decision matrices used for market segment evaluation and their visualizations, as well as explain how to use a generic segment evaluation plot in R. It will also address the criteria used to measure segment attractiveness and organizational competitiveness, along with the importance of identifying an ideal target segment in the market segmentation process.

The market segmentation team needs to determine the attractiveness of each segment based on specified criteria. The excerpt explains that to do this, the team assigns importance weights to each criterion and then rates each segment on a scale of 1 to 10 for each criterion. The team then multiplies the weight of each criterion by the rating for that criterion for each segment and adds up these weighted values to determine each segment's overall attractiveness.

The result of this calculation is used to plot each segment's attractiveness on a segment evaluation plot. The excerpt provides an example of such a plot, in which the x-axis represents how attractive the segment is to the organization and the y-axis represents how attractive the organization is to the segment. The size of each bubble on the plot represents some additional factor, such as profit potential.



The excerpt also notes that the same process is used to determine the organization's relative competitiveness for each segment, which is also plotted on the segment evaluation plot. By looking at the plot, the organization can identify the most attractive segments to focus on.

Implementaation of the Case studies

Name	Github Link
Balaji Kartheek	https://github.com/balajikartheek/Feynn_labs_Project_2
Atharva Kapile	https://github.com/atharvakap/FastFoodCaseStudy.git
Bhavneek Singh	https://github.com/blazingbhavneek/market_segmentation_case_study
Jasmit Singht	https://github.com/iamjasmit/Mcdonalds-case-study