Market segmentation

Analysis of the fast-food joint industry for marketing and business with McDonald's

By
Group - G

Raja Sai Nandhan Yadav Kataru

Anirudh JM

Sushrut Pandey

Harsha Lokesh

Code:- https://github.com/AnirudhJM24/MarketSegmentation

INTRODUCTION

Market segmentations are crucial for any business needs in order to make its business feasible and profit-driven in the future. Mcdonald's has applied business analysis and data-driven product development in order to remain sustainable and have great growth over the years. The need for market segmentation would include the change in prices of fast food products to target specific customer bases, customer tastes, and the development of products for better profit margins.

Step 1
Deciding (not) to segment

Implications of Committing to Market Segmentation
Understanding the implication of market segmentation is important before investing time and money into market segmentation. Market segmentation is a permanent and long-term generally and commitment, willingness to change and adapt to the situation is the most important criteria to move on. As Cahill says market segmentation is not recommended if there isn't any increase in sales, as this needs to be more profitable than marketing without it, the expenses of development and usage of the system needs to be looked into before the investment of time and money into the segmentation.

Implementation barriers
Barriers in market segmentation include senior management as they need to authorize the funds and implement the conclusion in the right way. Lack of consumer orientation, resistance to change and new ideas, lack of creative thinking, bad communication politics, etc, have been identified as preventing the successful implementation of market segmentation.
Financial resources and the inability to make structural changes are also major factors affecting an organization for effective implementation. Process-related barriers include not having clarified the objectives of the market segmentation exercise, bad planning, a lack of structure for the process, a lack of allocation of responsibilities, and time pressure that stands in the way of trying to find the best possible segmentation
In order to make it understandable and for usage, the use of graphical visualization will go a long way, if the barriers can not be removed the abandonment of segmentation is the best way to go in order to save money and time.

Step 2
Specifying the Ideal Target Segment

Important for user inputs of the data to not be limited to either a brief at the start of the process, development at the end. Users need to be involved in most stages, also at the technical aspects of market segmentation analysis.

Once committed to segmentation strategy in Step 1, the organisation/company has to improve to market segmentation analysis in Step 2. The organization must determine two sets of segment evaluation criteria.Knock-out criteria and attractiveness criteria are the main criterias to be considered before start of segmentation, the first is a non negotiating feature that is not upto the segmentation team as it is essential and cannot be tampered with and the other the criteria to determine the market segment selection.

Knock-out criteria
The organization members or owners must be satisfied with the members' needs. The capability to satisfy the organizations profits and consumers needs are important, members of the segment must be identifiable and part must be homogeneous as mentioned by Kotler.

Attractiveness criteria
Attractiveness across all criteria determines whether a market segment is selected as a target segment in Step 8 of market segmentation analysis.

Step 3
Data collection

Initial data:-

	yummy	convenient	spicy	fattening	greasy	fast	cheap	tasty	expensive	healthy	disgusting	Like	Age	VisitFrequency	Gender
0	No	Yes	No	Yes	No	Yes	Yes	No	Yes	No	No	-3	61	Every three months	Female
1	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	No	+2	51	Every three months	Female
2	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	+1	62	Every three months	Female
3	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	No	Yes	+4	69	Once a week	Female
4	No	Yes	No	Yes	Yes	Yes	Yes	No	No	Yes	No	+2	49	Once a month	Male

Segmentation variables
Market segmentation is extensively data driven and the need for useful and accurate data is important for correct results. In case of data driven in our case of Mcdonald’s data the dataset contains the specific menu items quality,frequency had by the customer, age, gender and likeness.

Segmentation Criteria
The term segmentation criterion relates to the nature of the information used for market segmentation. The decision on which segmentation criterion to use cannot easily be outsourced to either a consultant or a data analyst because it requires prior knowledge about the market. The most common segmentation criteria are geographic, socio-demographic, psychographic, and behavioral. If demographic segmentation works for your product or

service, then use demographic segmentation. Geographic segmentation will work because the product will only appeal to people in a particular region. Psychographic segmentation is appealing and more sophisticated than demographic or geographic segmentation, it is not better. Best is what works for the product or service at the least possible cost.

Choice of variables

The correct selection of variables in commonsense segmentation and data-driven segmentation is critical to the quality of the segmentation solution.

Sample data can be collected with surveys or data from internal sales and experimental studies.

Step 4

Exploring data

First look at data-

Exploratory data analysis cleans and, if necessary, pre-processes the data after it has been collected. This round of investigation also provides recommendations for the best algorithm for extracting useful market segments.

Data exploration, on a more technical level, aids in

(1) identifying the variables' measurement levels;

(2) investigating the univariate distributions of each variable; and

(3) assessing dependency patterns between variables.

Furthermore, data may need to be pre-processed and prepped before being utilised as input for various segmentation algorithms.The data exploration stage yields information on the suitability of various segmentation approaches for extracting market segments.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1453 entries, 0 to 1452
Data columns (total 15 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   yummy           1453 non-null   object
 1   convenient      1453 non-null   object
 2   spicy           1453 non-null   object
 3   fattening       1453 non-null   object
 4   greasy          1453 non-null   object
 5   fast            1453 non-null   object
 6   cheap           1453 non-null   object
 7   tasty           1453 non-null   object
 8   expensive       1453 non-null   object
 9   healthy         1453 non-null   object
 10  disgusting      1453 non-null   object
 11  Like            1453 non-null   object
 12  Age             1453 non-null   int64
 13  VisitFrequency  1453 non-null   object
 14  Gender          1453 non-null   object
dtypes: int64(1), object(14)
memory usage: 170.4+ KB
```

Data cleaning

Cleaning the data is the first step before beginning data analysis. This involves ensuring that all values have been appropriately recorded and that consistent labels for categorical variable levels have been utilised. The range of feasible values for many metric variables is known ahead of time. For example, one's age (in years) should be between 0 and 110. It's simple to see if the data has any unusual values, which could indicate problems in data gathering or data entry.

Summary after conversion of categorical data

	yummy	convenient	spicy	fattening	greasy	fast	cheap	tasty	expensive	healthy	disgusting	Like	Age	Gender
count	1453.000000	1453.000000	1453.000000	1453.000000	1453.000000	1453.000000	1453.000000	1453.000000	1453.000000	1453.000000	1453.000000	1453.000000	1453.000000	1453.000000
mean	0.552650	0.907777	0.093599	0.867171	0.526497	0.900206	0.598761	0.644184	0.357880	0.198899	0.242946	2.757743	44.604955	0.457674
std	0.497391	0.289440	0.291371	0.339506	0.499469	0.299828	0.490318	0.478925	0.479542	0.399309	0.429010	1.645749	14.221178	0.498377
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	18.000000	0.000000
25%	0.000000	1.000000	0.000000	1.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	33.000000	0.000000
50%	1.000000	1.000000	0.000000	1.000000	1.000000	1.000000	1.000000	1.000000	0.000000	0.000000	0.000000	3.000000	45.000000	0.000000
75%	1.000000	1.000000	0.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	0.000000	0.000000	4.000000	57.000000	1.000000
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	5.000000	71.000000	1.000000

Descriptive Analysis

It's easier to avoid misinterpretation of sophisticated analysis results if you're familiar with the data. Insights into the data can be gained through descriptive numerical and pictorial representations.

For descriptive analysis, statistical software packages provide a wide range of capabilities. With the command summary in R, we get a numeric summary of the data ().For numeric variables, this command returns the range, quartiles, and mean. The command returns frequency counts for categorical variables. For each variable, the command also provides the number of missing values. Histograms, box plots, and scatter plots are useful graphical approaches for numeric data. Frequency count bar charts are effective for visualising categorical variables. Mosaic plots show the relationship between numerous category variables.

Pre-possessing

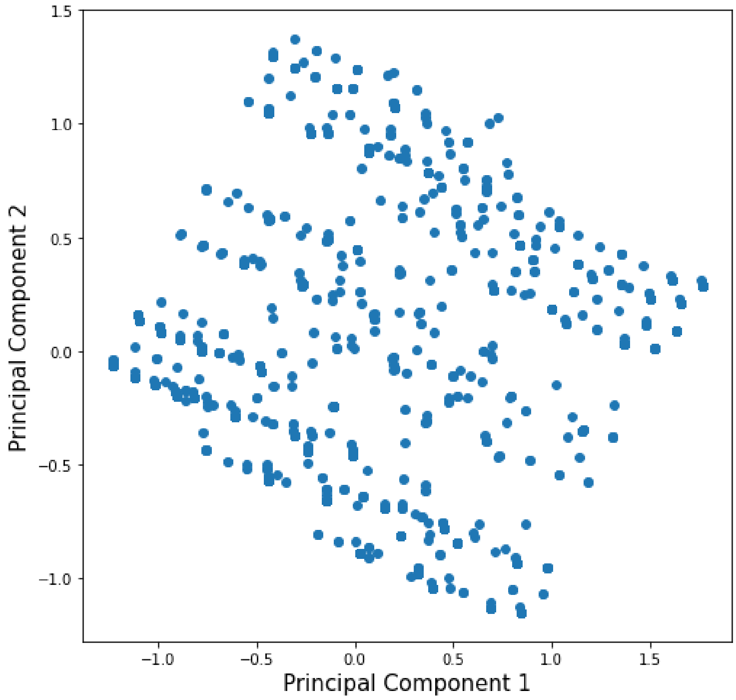
1) Categorical variables
   For categorical variables, two pre-processing approaches are commonly utilised. One is merging categorical variable levels before further analysis, and the other is transforming categorical variables to numeric variables if it makes sense. If the initial categories are too differentiated, merging levels of category variables can help (too many).
2) Numerical variables
   In distance-based methods of segment extraction, the range of values of a segmentation variable determines its relative influence. If one of the segmentation variables is binary (with values 0 or 1 indicating whether or not a customer views on the product of fast food), and a second variable indicates the expenditure in dollars per person per day (with values ranging from zero to $1000), a one-dollar difference in spend per person per day is weighted equally as the difference in liking to dine out or not. Factors can be standardised to balance the influence of segmentation variables on segmentation outcomes. Standardizing variables entails converting them into a scale that everyone can understand.

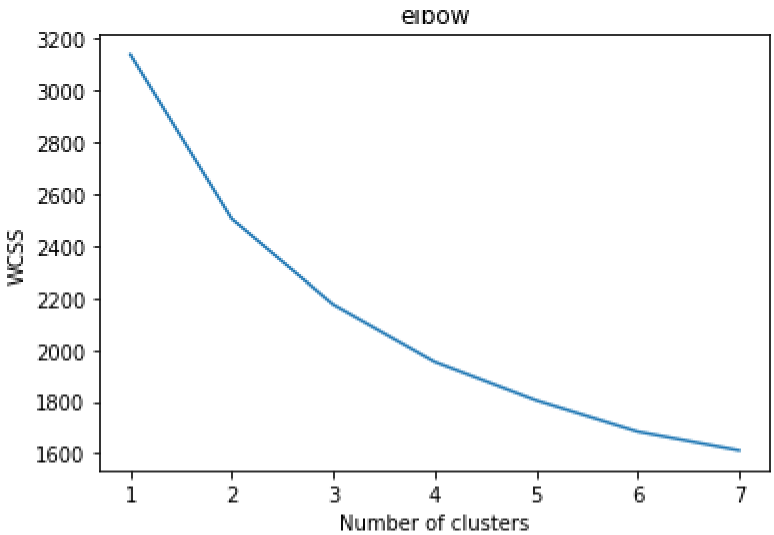
Principal component analysis

Principle components analysis (PCA) converts a multivariate data set with metric variables into a new data set with uncorrelated and importance-ordered variables called principal components. The most variability is contained in the first variable (principal component), the second principle component contains the second most variability, and so on. Because principal components analysis generates as many new variables as there were old ones, observations (consumers) retain their relative positions to one another after transformation, and the dimensionality of the new data set remains the same. The data space is essentially unchanged in principal components analysis, but it is viewed from a different perspective.

Prior to creating market segments from consumer data, principal components analysis may be used to reduce the number of segmentation factors. This concept appeals to me since more variables increase the intricacy of the problem that the segment extraction technique must deal with, making extraction more difficult and requiring larger sample sizes. The early segmentation literature also advised reducing dimensionality by selecting only a small number of primary components, however this has now been proven to be extremely problematic.

<matplotlib.collections.PathCollection at 0x7fb7f6ee6650>





K-means and k-centroid clustering

K-means clustering is the most often used partitioning method. A lot of algorithms are available inside this technique.The R function kmeans() implements Forgy's (1965), Hartigan and Wong's (1979), Lloyd's (1982), and MacQueen's (1982) algorithms (1967). The squared Euclidean distance is used in these algorithms. The R package flexclust provides a generalisation to various distance measurements, often known as k-centroid clustering.

In a data set, let X = x1,..., xn represents a set of observations (consumers). Customers assigned to the same market segment are as similar as feasible, while consumers belonging to different market segments are as diverse as possible, according to partitioning clustering methodologies. Many partitioning clustering techniques refer to the centroid as the representation of a market segment. The centroid of the k-means method based on the squared Euclidean distance is made up of the column-wise mean values across all market segment members. In the data set, there are rows of observations (consumers) and columns of variables (behavioural data or survey responses).

Step 5
Extracting data

Grouping Consumers
Data-driven market segmentation analysis is exploratory by nature. Consumer data sets are typically not well structured. Consumers come in all shapes and forms; a two-dimensional plot of consumers' product preferences typically does not contain clear groups of consumers. Rather, consumer preferences are spread across the entire plot. 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. The result of a market segmentation analysis, therefore, 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.

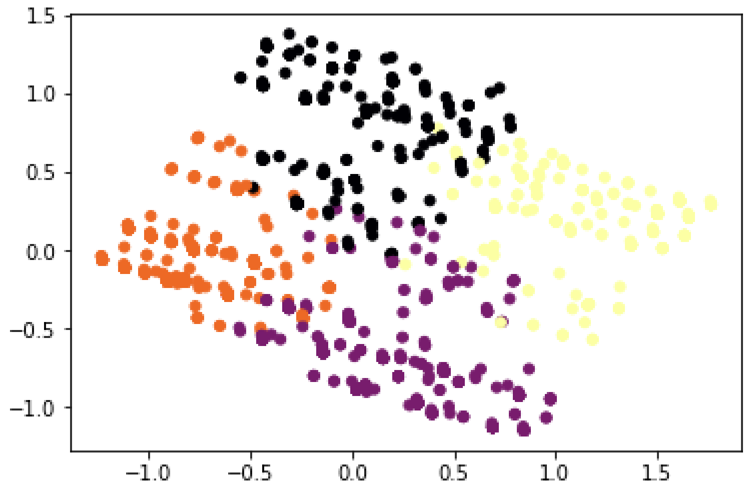
The segmentation Processes are broadly classified into two types:
1) Distance-based
2) Model-based

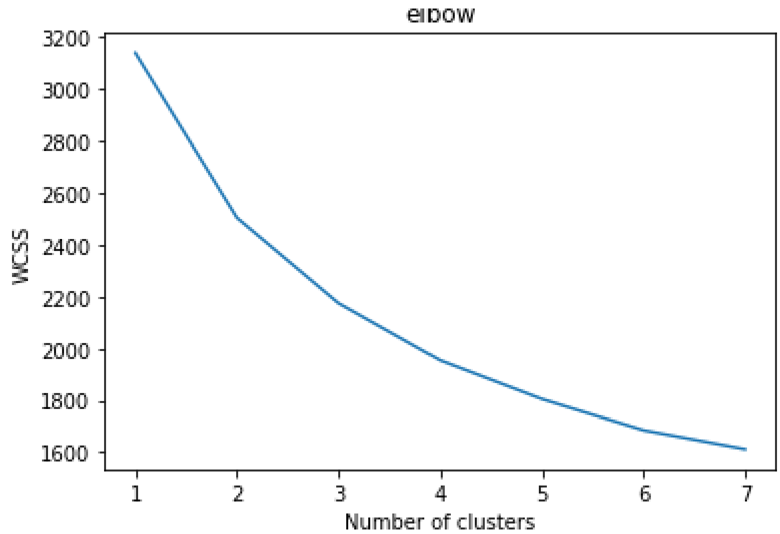
1) Distance-based
   Distance-based methods are further divided into many methods such as
   1. Distance measures
   2. Hierarchical measures
   3. Partitioning method
   4. Hybrid approach
2) Model-based
   Distributions include following
   1. Finite Mixtures of Distributions-
      1. Normal distribution
      2. Binary distribution
   2. Finite mixtures of regressions-
   3. Algorithms with Integrated variable selection
      1. Biclustering algorithms
      2. Variable reduction: factor-cluster analysis
      3. Variable selection procedure for clustering binary data (VSBD)

K-means and k-centroid clustering

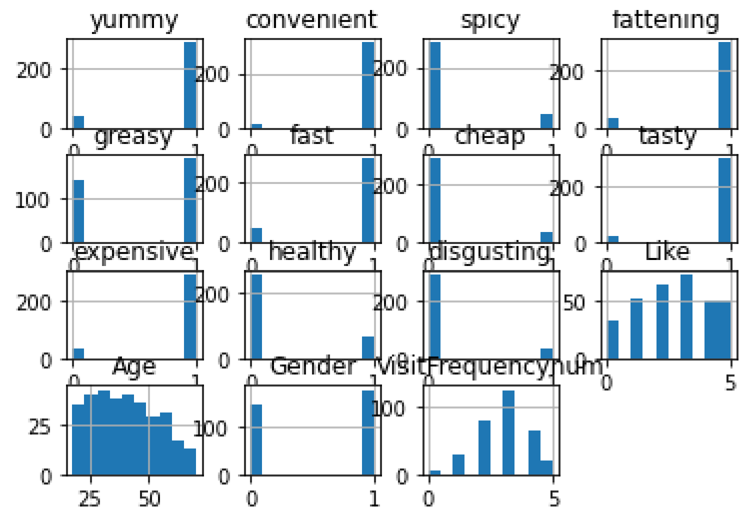
K-means clustering is the most often used partitioning method. A lot of algorithms are available inside this technique.The R function kmeans() implements Forgy's (1965), Hartigan and Wong's (1979), Lloyd's (1982), and MacQueen's (1982) algorithms (1967). The squared Euclidean distance is used in these algorithms. The R package flexclust provides a generalisation to various distance measurements, often known as k-centroid clustering.

In a data set, let X = x1,..., xn represents a set of observations (consumers). Customers assigned to the same market segment are as similar as feasible, while consumers belonging to different market segments are as diverse as possible, according to partitioning clustering methodologies. Many partitioning clustering techniques refer to the centroid as the representation of a market segment. The centroid of the k-means method based on the squared Euclidean distance is made up of the column-wise mean values across all market segment members. In the data set, there are rows of observations (consumers) and columns of variables (behavioural data or survey responses).





Graphical visualization



Cluster indices

Because market segmentation analysis is exploratory, data analysts require assistance in making some of the most important judgments, such as the number of market segments to extract.

The most popular method for gaining such assistance is to use so-called cluster indices.Cluster indices shed light on specific components of a market segmentation approach. The type of insight obtained is determined by the cluster index chosen.

Internal cluster indices and exterior cluster indices are the two types of cluster indices that are commonly used.Internal cluster indices are produced on the basis of a single market segmentation solution and provide suggestions based on the information contained in that solution.The sum of all distances between pairs of segment members is an example of an internal cluster index. The lower this value is, the more members of the same segment are similar. Users are drawn to segments with members who are similar.

Gorge plot

Examining the distances between each consumer and all segment representatives is an easy way to determine how well segments are divided.Let dih indicate the distance between customer I and the segment representative (centroid, cluster centre) h. Then,

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

with the hyper parameter determining how variations in distance translate into differences in similarity, can be understood as the similarity of customer I to the representative of segment h. These parallels range from 0 to 1 and add up to 1 for each consumer I across all segment representatives h, h = 1,...,k.

Global Stability Analysis

Resampling methods are an alternative to distance and model-based segment extraction techniques for data structure research. The stability of a market segmentation solution can be determined using resampling methods. Several fresh data sets are created using resampling methods, and a number of segmentation solutions are retrieved to evaluate the global stability of any particular segmentation solution.

Step 6

Profiling Segments

Identifying Key Characteristics of Market Segments

The purpose of the profiling step is to learn about the market segments that were discovered during the extraction process. Only when data-driven market segmentation is used does profiling become necessary. The profiles of the segments are predefined in commonsense segmentation. If, for example, age is employed as a segmentation variable in commonsense segmentation, the resulting segments will undoubtedly be age groups. As a result, while using commonsense segmentation, Step 6 is not required.

In the case of data-driven segmentation, the scenario is completely different: segmentation solution users may have chosen to extract categories based on consumer benefits sought. The distinguishing qualities of the resulting market categories, however, are unknown until the data has been analysed. The goal of profiling is to identify these defining characteristics of market segments in relation to segmentation variables. Profiling entails describing each market segment separately as well as in reference to other market segments. When questioned about their vacation activities, the majority of winter tourists in Austria say they are going alpine skiing. Alpine skiing may characterise a market segment, but it may not distinguish it from other market segments.

Segment Profiling with Visualisations

Although data visualisation using graphics is a vital aspect of statistical data analysis, neither the highly simplified nor the very complicated tabular representations commonly used to demonstrate market segmentation solutions make extensive use of visuals (Tufte 1983, 1997; Cleveland 1993; Chen et al. 2008; Wilkinson 2005; Kastellec and Leoni 2007). In exploratory statistical analysis (such as cluster analysis), graphics are very useful because they reveal the complicated relationships between variables. Furthermore, in an age of enormous data, visualisation provides a straightforward approach to track changes over time. Visualization tools are recommended by both McDonald and Dunbar (2012) and Lilien and Rangaswamy (2003) to make market segmentation analysis data easier to interpret.

In the data-driven market segmentation process, visualisations are important for inspecting one or more segments in detail for each segmentation solution. The comprehension of segment profiles is made easier by statistical graphs. They also make evaluating the utility of a market segmentation approach more easier. When it comes to data segmentation, there are always a lot of different options. The choice of one of the various options is a crucial one. This task is made easier for the data analyst and user by using visualisations of solutions.

1) Identifying Defining Characteristics of Market Segments
2) Assessing Segment Separation

Step 7
Describing Segments

Developing a Complete Picture of Market Segments
Understanding differences in segmentation variables across market segments is the goal of segment profiling. Early in the market segmentation study process, segmentation variables are chosen conceptually in Step 2 (specifying the ideal target segment) and empirically in Step 3. (collecting data).
Market segments are extracted from empirical data using segmentation characteristics. The seventh phase is very similar to the profiling step. The only distinction is that the factors being examined were not used to identify market groups. Rather, market segments are described in Step 7 utilising extra information about segment members that is accessible. If committing to a target segment is like getting married, profiling and characterising market segments is like going on several dates to get to know the potential spouse as well as possible in order to give the marriage the best possible opportunity and avoid unpleasant shocks later on. Segment... should be further characterised and characterised by crossing them with all other characteristics, such as psychographic, demographic, and socio-economic variables, media exposure, and specific product and brand attitudes or assessments, as stated by van-Raaij/Verhallen.

Using Visualisations to Describe Market Segments
Differences in descriptor variables can be seen using a variety of visualisations. We'll go over two basic ways for nominal and ordinal descriptor variables (such gender, education level,

and country of origin), as well as metric descriptor variables (such as age, number of nights at the tourist destinations, money spent on accommodation).

1) Nominal and Ordinal Descriptor Variables
2) Metric Descriptor Variables

Testing for Segment Differences in Descriptor Variables

To formally test for differences in descriptor variables across market sectors, simple statistical tests can be utilised. Running a series of independent tests for each variable of interest is the simplest technique to test for differences. Segment membership, or the assignment of each consumer to one market segment, is the result of the segment extraction phase. Segment membership is a nominal variable that can be treated like any other. It represents the segmentation variables' nominal summary statistics. As a result, any test for a nominal variable's relationship with another variable will suffice.

Predicting Segments from Descriptor Variables

Another technique to learn about market segments is to use descriptor variables to predict segment membership. To do so, we utilise a regression model with descriptor variables as independent variables and segment membership as a categorical dependent variable. For classification, we can use methods created in statistics, and for supervised learning, we can utilise methods developed in machine learning.

Step 10

Selecting the Target Segment(s)

The Targeting Decision

This is where the rubber meets the road in Step 8. The key choice has been made: which of the many potential market segments will be targeted? Market segmentation is a marketing strategy. The choice of one or more target segments is a long-term decision that has a substantial impact on an organization's future performance. It's time to buy a ring, pop the question, and commit after all of the flirting and dating.

Market Segment Evaluation

The use of a decision matrix to visualise relative segment attractiveness and relative organisational competitiveness for each market segment is recommended in most works on target market selection (e.g., McDonald and Dunbar 1995; Lilien and Rangaswamy 2003).

Many names have been used to describe decision matrices in the past, including Boston matrix (McDonald and Dunbar 1995; Dibb and Simkin 2008) because this type of matrix was first proposed by the Boston Consulting Group; General Electric / McKinsey matrix because this extended version of the matrix was jointly developed by General Electric and McKinsey; directional policy matrix (McDonald and Dunbar 1995; Dibb and Simkin 2008); McKinsey