***Summary of the Book***

***Market Segmentation Analysis***

***Understanding it, Doing it,***

***And Making It Useful***

***Team Members:***

***S Deepika***

***Mohammad Ammaruddin***

***Kunal Aggarwal***

***Gouri Rajendra Uplenchwar***

***Krishnendu Barman***

***Step 1:Implications of Committing to Market Segmentation***

The key implication is that the needs to commit to the segmentation strategy on the long term. . The commitment to market segmentation goes hand in hand with the willingness and ability of the to make substantial changes (McDonald and Dunbar 1995) and investments.

Modification of existing products, changes in pricing and distribution channels used to sell the product, as well as all communications with the market. These changes, in turn, are likely to influence the internal structure of the , which may need to be adjusted in view of, for example, targeting a handful of different market segments. Croft (1994) recommends that – to maximize the benefits of market segmentation – s need to organize around (p. 66) market, Step 1: Deciding (not) to Segment segments, rather than organizing around products.

3.2 Implementation Barriers:

The first group of barriers relates to senior management. Lack of leadership, pro-active championing, commitment and involvement in the market segmentation process by senior leadership undermines the success of market segmentation. As McDonald and Dunbar (1995, p. 158) state: There can be no doubt that unless the chief executive sees the need for a segmentation review.

Senior management can also prevent market segmentation to be successfully implemented by not making enough resources available, either for the initial market segmentation analysis itself, or for the long-term implementation of a market segmentation strategy. A second group of barriers relates to organisational culture. Lack of market or consumer orientation, resistance to change and new ideas, lack of creative thinking, bad communication and lack of sharing of information and insights across organisational units, short-term thinking,

Another potential problem is lack of training.

Another obstacle may be objective restrictions faced by the , including lack of financial resources, or the inability to make the structural changes required

 A company with limited resources needs to pick only the best opportunities to pursue. Process-related barriers include not having clarified the objectives of the market segmentation exercise, lack of planning or bad planning, a lack of structured processes to guide the team through all steps of the market segmentation process, a lack of allocation of responsibilities, and time pressure that stands in the way of trying to find the best possible segmentation outcome.

3 Step 1 Checklist

This first checklist includes not only tasks, but also a series of questions which, if not answered in the affirmative, serve as knock-out criteria. For example: if an is not market-oriented, even the finest of market segmentation analyses cannot be successfully implemented.

# *Step 2: Specifying the Ideal Target Segment*

## Abstract

Market segmentation analysis is driven primarily by the desire of an to better cater to a part of the market and, in so doing, secure a competitive advantage. At the end of the segmentation analysis, the needs to select one or more target segments.

After having committed to investigating the value of a segmentation strategy in Step [1](https://doi.org/10.1007/978-981-10-8818-6_1), the has to make a major contribution to market segmentation analysis in Step 2. While this contribution is conceptual in nature, it guides many of the following steps, most critically Step [3](https://doi.org/10.1007/978-981-10-8818-6_3) (data collection) and Step [8](https://doi.org/10.1007/978-981-10-8818-6_8) (selecting one or more target segments). In Step 2 the must determine two sets of segment evaluation criteria. One set of evaluation criteria can be referred to as *knock-out criteria*. These criteria are the essential, non-negotiable features of segments that the would consider targeting. The second set of evaluation criteria can be referred to as *attractiveness criteria*. These criteria are used to evaluate the relative attractiveness of the remaining market segments – those in compliance with the knock-out criteria.

Where knock-out criteria automatically eliminate some of the available market segments, attractiveness criteria are first negotiated by the team, and then applied to determine the overall relative attractiveness of each market segment in Step [8](https://doi.org/10.1007/978-981-10-8818-6_8).

## 2 Knock-Out Criteria

Knock-out criteria are used to determine if market segments resulting from the market segmentation analysis qualify to be assessed using segment attractiveness criteria. The first set of such criteria was suggested by Kotler ([1994](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_4#ref-CR6)) and includes substantiality, measurability and accessibility (Tynan and Drayton [1987](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_4#ref-CR18)). Kotler himself and a number of other authors have since recommended additional criteria that fall into the knock-out criterion category (Wedel and Kamakura [2000](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_4#ref-CR19); Lilien and Rangaswamy [2003](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_4#ref-CR8); McDonald and Dunbar [2012](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_4#ref-CR11)):

* The segment must be **homogeneous**; members of the segment must be similar to one another.
* The segment must be **distinct**; members of the segment must be distinctly different from members of other segments.
* The segment must be **large enough**; the segment must contain enough consumers to make it worthwhile to spend extra money on customizing the marketing mix for them.
* The segment must be **matching** the strengths of the organization; the must have the capability to satisfy segment members’ needs.
* Members of the segment must be **identifiable**; it must be possible to spot them in the marketplace.
* The segment must be **reachable**; there has to be a way to get in touch with members of the segment in order to make the customized marketing mix accessible to them.

Knock-out criteria must be understood by senior management, the segmentation team, and the advisory committee. Most of them do not require further specification, but some do. For example, while size is non-negotiable, the exact minimum viable target segment size needs to be specified.

## 3 Attractiveness Criteria

Attractiveness criteria are not binary in nature. Segments are not assessed as either complying or not complying with attractiveness criteria. Rather, each market segment is rated; it can be more or less attractive with respect to a specific criterion. The attractiveness across all criteria determines whether a market segment is selected as a target segment in Step [8](https://doi.org/10.1007/978-981-10-8818-6_8) of market segmentation analysis.

## 4 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.

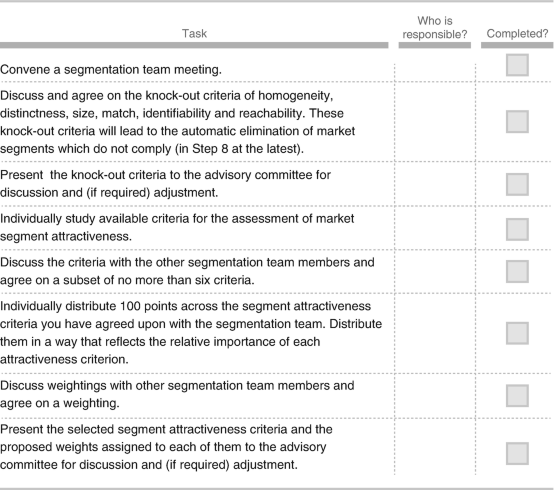
 The segment attractiveness and organisational competitiveness values are determined by the segmentation team. This is necessary because there is no standard set of criteria that could be used by all organizations.

Factors which constitute both segment attractiveness and organisational competitiveness need to be negotiated and agreed upon. To achieve this, a large number of possible criteria has to be investigated before agreement.

There are at least two good reasons to include in this process representatives from a wide range of organisational units. First, each organisational unit has a different perspective on the business of the organisation. As a consequence, members of these units bring different positions to the deliberations. Secondly, if the segmentation strategy is implemented, it will affect every single unit of the organization. Consequently, all units are key stakeholders of market segmentation analysis.

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 organization compared to the other criteria.

5 Step 2 Checklist



# *Step 3: Collecting Data*

## Abstract

The outcome of a market segmentation analysis is only as good as the data upon which it is based.

## 1 Segmentation Variables

Empirical data forms the basis of both commonsense and data-driven market segmentation. Empirical data is used to identify or create market segments and – later in the process – describe these segments in detail.

*segmentation variable* to refer to the variable in the empirical data used in commonsense segmentation to split the sample into market segments. In commonsense segmentation, the segmentation variable is typically one single characteristic of the consumers in the sample. This case is illustrated in Table [5.1](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#Tab1). Each row in this table represents one consumer, each variable represents one characteristic of that consumer. An entry of 1 in the data set indicates that the consumer has that characteristic. An entry of 0 indicates that the consumer does not have that characteristic.

All the other personal characteristics available in the data – in this case: age, the number of vacations taken, and information about five benefits people seek or do not seek when they go on vacation – serve as so-called *descriptor variables* . They are used to describe the segments in detail. Describing segments is critical to being able to develop an effective marketing mix targeting the segment. Typical descriptor variables include socio-demographics, but also information about media behavior, allowing marketers to reach their target segment with communication messages.

The difference between commonsense and data-driven market segmentation is that data-driven market segmentation is based not on one, but on multiple segmentation variables. These segmentation variables serve as the starting point for identifying naturally existing, or artificially creating market segments useful to the . Data quality is critical to both (1) assigning each person in the sample to the correct market segment, and (2) being able to correctly describe the segments. The correct description, in turn, makes it possible to develop a customized product, determine the most appropriate pricing strategy, select the best distribution channel, and the most effective communication channel for advertising and promotion.

2 Segmentation Criteria

The term *segmentation criterion* is used here in a broader sense than the term segmentation variable. The term segmentation variable refers to one measured value, for example, one item in a survey, or one observed expenditure category. The term segmentation criterion relates to the nature of the information used for market segmentation. It can also relate to one specific construct, such as benefits sought.

### 2.3 Psychographic Segmentation

When people are grouped according to psychological criteria, such as their beliefs, interests, preferences, aspirations, or benefits sought when purchasing a product, the term psychographic segmentation is used. Most psychographic segmentation studies use a number of segmentation variables, for example: a number of different travel motives, a number of perceived risks when going on vacation.

The psychographic approach has the advantage that it is generally more reflective of the underlying reasons for differences in consumer behavior. For example, tourists whose primary motivation to go on vacation is to learn about other cultures, have a high likelihood of undertaking a cultural holiday at a destination that has ample cultural treasures for them to explore.

### 2.4 Behavioral Segmentation

 A wide range of possible behaviors’ can be used for this purpose, including prior experience with the product, frequency of purchase, amount spent on purchasing the product on each occasion (or across multiple purchase occasions), and information search behavior. In a comparison of different segmentation criteria used as segmentation variables, behaviors reported by tourists emerged as superior to geographic variables.

The key advantage of behavioral approaches is that – if based on actual behavior rather than stated behavior or stated intended behavior – the very behavior of interest is used as the basis of segment extraction.

But behavioral data is not always readily available, especially if the aim is to include in the segmentation analysis potential customers who have not previously purchased the product, rather than limiting oneself to the study of existing customers of the organization.

3 Data from Survey Studies

 survey data – as opposed to data obtained from observing actual behavior – can be contaminated by a wide range of biases. Such biases can, in turn, negatively affect the quality of solutions derived from market segmentation analysis. A few key aspects that need to be considered when using survey data are discussed below.

### 3.1 Choice of Variables

Carefully selecting the variables that are included as segmentation variable in commonsense segmentation, or as segmentation variables in data-driven segmentation, is critical to the quality of the market segmentation solution.

In data-driven segmentation, all variables relevant to the construct captured by the segmentation criterion need to be included. At the same time, unnecessary variables must be avoided.

 Unnecessary variables included as segmentation variables divert the attention of the segment extraction algorithm away from information critical to the extraction of optimal market segments.

 Redundant items are particularly problematic in the context of market segmentation analysis because they interfere substantially with most segment extraction algorithms’ ability to identify correct market segmentation solutions

### 3.2 Response Options

Answer options provided to respondents in surveys determine the scale of the data available for subsequent analyses. Because many data analytic techniques are based on distance measures , not all survey response options are equally suitable for segmentation analysis.

Options allowing respondents to answer in only one of two ways, generate *binary* or *dichotomous data*. Such responses can be represented in a data set by 0s and 1s. The distance between 0 and 1 is clearly defined and, as such, poses no difficulties for subsequent segmentation analysis.

### 3.3 Response Styles

A wide range of response styles manifest in survey answers, including respondents’ tendencies to use extreme answer options (STRONGLY AGREE, STRONGLY DISAGREE), to use the midpoint (NEITHER AGREE NOR DISAGREE), and to agree with all statements. Response styles affect segmentation results because commonly used segment extraction algorithms cannot differentiate between a data entry reflecting the respondent’s belief from a data entry reflecting both a respondent’s belief and a response style.

### 3.4 Sample Size

The market segmentation problem in this figure is extremely simple because only two segmentation variables are used. Yet, when the sample size is insufficient (left plot), it is impossible to determine which the correct number of market segments is. If the sample size is sufficient, however (right plot) it is very easy to determine the number and nature of segments in the data set.

The outcome of a market segmentation analysis is only as good as the data upon which it is based. This chapter discusses a range of alternative sources of data that can serve as input for extracting market segments. Key potential dangers associated with each of those sources are discussed. A checklist summarises a number of questions that may assist in ensuring that data of the highest quality is being collected.

## 1 Segmentation Variables

Empirical data forms the basis of both commonsense and data-driven market segmentation. Empirical data is used to identify or create market segments and – later in the process – describe these segments in detail.

Throughout this book we use the term *segmentation variable* to refer to the variable in the empirical data used in commonsense segmentation to split the sample into market segments. In commonsense segmentation, the segmentation variable is typically one single characteristic of the consumers in the sample.. Each row in this table represents one consumer, each variable represents one characteristic of that consumer. An entry of 1 in the data set indicates that the consumer has that characteristic. An entry of 0 indicates that the consumer does not have that characteristic. The commonsense segmentation illustrated in Table [5.1](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#Tab1) uses gender as the segmentation variable. Market segments are created by simply splitting the sample using this segmentation variable into a segment of women and a segment of men.

**Table 5.1 Gender as a possible segmentation variable in commonsense market segmentation**

All the other personal characteristics available in the data – in this case: age, the number of vacations taken, and information about five benefits people seek or do not seek when they go on vacation – serve as so-called *descriptor variables* . They are used to describe the segments in detail. Describing segments is critical to being able to develop an effective marketing mix targeting the segment. Typical descriptor variables include socio-demographics, but also information about media behavior, allowing marketers to reach their target segment with communication messages.

The difference between commonsense and data-driven market segmentation is that data-driven market segmentation is based not on one, but on multiple segmentation variables. These segmentation variables serve as the starting point for identifying naturally existing, or artificially creating market segments useful to the organization..

**Table 5.2 Segmentation variables in data-driven market segmentation**

In the data-driven case we may, for example, want to extract market segments of tourists who do not necessarily have gender in common, but rather share a common set of benefits they seek when going on vacation. Sorting the data from Table [5.1](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#Tab1) using this set of segmentation variables reveals one segment (shown in the first three rows) characterized by seeking relaxation, culture and meeting people, but not interested in action and exploring. In this case, the benefits sought represent the segmentation variables. The socio-demographic variables, gender, age, and the number of vacations undertaken per annum serve as descriptor variables.

These two simple examples illustrate how critical the quality of empirical data is for developing a valid segmentation solution. When commonsense segments are extracted – even if the nature of the segments is known in advance – data quality is critical to both (1) assigning each person in the sample to the correct market segment, and (2) being able to correctly describe the segments. The correct description, in turn, makes it possible to develop a customized product, determine the most appropriate pricing strategy, select the best distribution channel, and the most effective communication channel for advertising and promotion.

The same holds for data-driven market segmentation where data quality determines the quality of the extracted data-driven market segments, and the quality of the descriptions of the resulting segments. Good market segmentation analysis requires good empirical data .

Empirical data for segmentation studies can come from a range of sources: from survey studies; from observations such as scanner data where purchases are recorded and, frequently, are linked to an individual customer’s long-term purchase history via loyalty programs; or from experimental studies. Optimally, data used in segmentation studies should reflect consumer behavior . Survey data – although it arguably represents the most common source of data for market segmentation studies – can be unreliable in reflecting behavior, especially when the behavior of interest is socially desirable, such as donating money to a charity or behaving in an environmentally friendly way (Karlsson and Dolnicar [2016](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR30)). Surveys should therefore not be seen as the default source of data for market segmentation studies. Rather, a range of possible sources should be explored. The source that delivers data most closely reflecting actual consumer behavior is preferable.

## 2 Segmentation Criteria

Long before segments are extracted, and long before data for segment extraction is collected, the organization must make an important decision: it must choose which segmentation criterion to use (Tynan and Drayton [1987](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR48)). The term *segmentation criterion* is used here in a broader sense than the term segmentation variable. The term segmentation variable refers to one measured value, for example, one item in a survey, or one observed expenditure category. The term segmentation criterion relates to the nature of the information used for market segmentation. It can also relate to one specific construct, such as benefits sought.

The decision 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.

Bock and Uncles ([2002](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR2)) argue that the following differences between consumers are the most relevant in terms of market segmentation: profitability, bargaining power, preferences for benefits or products, barriers to choice and consumer interaction effects. With so many different segmentation criteria available, which is the best to use? As Hoek et al. ([1996](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR28)) note, few guidelines as to the most appropriate base to use in a given marketing context exist (p. 26). Generally, the recommendation is to use the simplest possible approach. Cahill ([2006](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR6)) states this very clearly in his book on lifestyle segmentation (p. 159): Do the least you can. If demographic segmentation will work for your product or service, then use demographic segmentation. If geographic segmentation will work because your product will only appeal to people in a certain region, then use it. Just because psychographic segmentation is sexier and more sophisticated than demographic or geographic segmentation does not make it better. Better is what works for your product or service at the least possible cost.

### 2.1 Geographic Segmentation

Geographic information is seen as the original segmentation criterion used for the purpose of market segmentation (Lewis et al. [1995](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR32); Tynan and Drayton [1987](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR48)). Typically – when geographic segmentation is used – the consumer’s location of residence serves as the only criterion to form market segments. For example: if the national tourism organization of Austria wants to attract tourists from neighboring countries, it needs to use a number of different languages: Italian, German, Slovenian, Hungarian, Czech. Language differences across countries represent a very pragmatic reason for treating tourists from different neighboring countries as different segments. Interesting examples are also provided by global companies such as Amazon selling its Kindle online: one common web page is used for the description of the base product, then customers are asked to indicate their country of residence and country specific additional information is provided. IKEA offers a similar product range worldwide, yet slight differences in offers, pricing as well as the option to purchase online exist in dependence of the customer’s geographic location.

The key advantage of geographic segmentation is that each consumer can easily be assigned to a geographic units.

The key disadvantage is that living in the same country or area does not necessarily mean that people share other characteristics relevant to marketers, such as benefits they seek when purchasing a product. Even in the case of luxury suburbs, it is more likely that socio-demographic criteria are the reason for both similar choice of suburb to live in and similar car preferences. The typical case is best illustrated using tourism: people from the same country of origin are likely to have a wide range of different ideal holidays, depending on whether they are single or travel as a family, whether they are into sports or culture.

Despite the potential shortcomings of using geographic information as the segmentation variable, the location aspect has experienced a revival in international market segmentation studies aiming to extract market segments across geographic boundaries. Such an approach is challenging because the segmentation variable(s) must be meaningful across all the included geographic regions, and because of the known biases that can occur if surveys are completed by respondents from different cultural backgrounds (Steenkamp and Ter Hofstede [2002](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR45)). An example of such an international market segmentation study is provided by Haverila ([2013](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR26)) who extracted market segments of mobile phone users among young customers across national borders.

### 2.2 Socio-Demographic Segmentation

Typical socio-demographic segmentation criteria include age, gender, income and education. Socio-demographic segments can be very useful in some industries. For example: luxury goods (associated with high income), cosmetics (associated with gender; even in times where men are targeted, the female and male segments are treated distinctly differently), baby products (associated with gender), retirement villages (associated with age), tourism resort products (associated with having small children or not).

As is the case with geographic segmentation, socio-demographic segmentation criteria have the advantage that segment membership can easily be determined for every consumer. In some instances, the socio-demographic criterion may also offer an explanation for specific product preferences (having children, for example, is the actual reason that families choose a family vacation village where previously, as a couple, their vacation choice may have been entirely different). But in many instances, the socio-demographic criterion is not the *cause* for product preferences, thus not providing sufficient market insight for optimal segmentation decisions. Haley ([1985](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR25)) estimates that demographics explain about 5% of the variance in consumer behavior. Yankelovich and Meer ([2006](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR49)) argue that socio-demographics do not represent a strong basis for market segmentation, suggesting that values, tastes and preferences are more useful because they are more influential in terms of consumers’ buying decisions.

### 2.3 Psychographic Segmentation

When people are grouped according to psychological criteria, such as their beliefs, interests, preferences, aspirations, or benefits sought when purchasing a product, the term psychographic segmentation is used. Haley ([1985](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR25)) explains that the word psychographics was intended as an umbrella term to cover all measures of the mind (p. 7). Benefit segmentation, which Haley ([1968](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR24)) is credited for, is arguably the most popular kind of psychographic segmentation. Lifestyle segmentation is another popular psychographic segmentation approach (Cahill [2006](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR6)); it is based on people’s activities, opinions and interests.

Psychographic criteria are, by nature, more complex than geographic or socio-demographic criteria because it is difficult to find a single characteristic of a person that will provide insight into the psychographic dimension of interest. As a consequence, most psychographic segmentation studies use a number of segmentation variables, for example: a number of different travel motives, a number of perceived risks when going on vacation.

The psychographic approach has the advantage that it is generally more reflective of the underlying reasons for differences in consumer behavior. For example, tourists whose primary motivation to go on vacation is to learn about other cultures, have a high likelihood of undertaking a cultural holiday at a destination that has ample cultural treasures for them to explore. Not surprisingly, therefore, travel motives have been frequently used as the basis for data-driven market segmentation in tourism (Bieger and Laesser [2002](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR1); Laesser et al. [2006](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR31); Boksberger and Laesser [2009](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR4)). The disadvantage of the psychographic approach is the increased complexity of determining segment memberships for consumers. Also, the power of the psychographic approach depends heavily on the reliability and validity of the empirical measures used to capture the psychographic dimensions of interest.

### 2.4 Behavioral Segmentation

Another approach to segment extraction is to search directly for similarities in behavior or reported behavior. A wide range of possible behaviors can be used for this purpose, including prior experience with the product, frequency of purchase, amount spent on purchasing the product on each occasion (or across multiple purchase occasions), and information search behavior. In a comparison of different segmentation criteria used as segmentation variables, behaviors reported by tourists emerged as superior to geographic variables (Moscardo et al. [2001](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR34)).

The key advantage of behavioral approaches is that – if based on actual behavior rather than stated behavior or stated intended behavior – the very behavior of interest is used as the basis of segment extraction. As such, behavioral segmentation groups people by the similarity which matters most. Examples of such segmentation analyses are provided by Tsai and Chiu ([2004](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR47)) who use actual expenses of consumers as segmentation variables, and Heilman and Bowman ([2002](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR27)) who use actual purchase data across product categories. Brand choice behavior over time has also been used as segmentation variable by several authors (Poulsen [1990](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR39); Bockenholt and Langeheine [1996](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR3); Ramaswamy [1997](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR41), see also Section [7.3.3](https://doi.org/10.1007/978-981-10-8818-6_7#Sec24)). Using behavioral data also avoids the need for the development of valid measures for psychological constructs.

But behavioral data is not always readily available, especially if the aim is to include in the segmentation analysis potential customers who have not previously purchased the product, rather than limiting oneself to the study of existing customers of the organization.

## 3 Data from Survey Studies

Most market segmentation analyses are based on survey data. Survey data is cheap and easy to collect, making it a feasible approach for any organization. But survey data – as opposed to data obtained from observing actual behavior – can be contaminated by a wide range of biases. Such biases can, in turn, negatively affect the quality of solutions derived from market segmentation analysis. A few key aspects that need to be considered when using survey data are discussed below.

### 3.1 Choice of Variables

Carefully selecting the variables that are included as segmentation variable in commonsense segmentation, or as segmentation variables in data-driven segmentation, is critical to the quality of the market segmentation solution.

In data-driven segmentation, all variables relevant to the construct captured by the segmentation criterion need to be included. At the same time, unnecessary variables must be avoided. Including unnecessary variables can make questionnaires long and tedious for respondents, which, in turn, causes respondent fatigue . Fatigued respondents tend to provide responses of lower quality (Johnson et al. [1990](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR29); Dolnicar and Rossiter [2008](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR17)). Including unnecessary variables also increases the dimensionality of the segmentation problem without adding relevant information, making the task of extracting market segments unnecessarily difficult for any data analytic technique. The issue of the appropriate ratio of the number of variables and the available sample is discussed later in this chapter. Unnecessary variables included as segmentation variables divert the attention of the segment extraction algorithm away from information critical to the extraction of optimal market segments. Such variables are referred to as *noisy variables* or *masking variables* and have been repeatedly shown to prevent algorithms from identifying the correct segmentation solution (Brusco [2004](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR5); Carmone et al. [1999](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR7); DeSarbo et al. [1984](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR11); DeSarbo and Mahajan [1984](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR10); Milligan [1980](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR33)).

Noisy variables do not contribute any information necessary for the identification of the correct market segments. Instead, their presence makes it more difficult for the algorithm to extract the correct solution. Noisy variables can result from not carefully developing survey questions, or from not carefully selecting segmentation variables from among the available survey items. The problem of noisy variables negatively affecting the segmentation solution can be avoided at the data collection and the variable selection stage of market segmentation analysis.

The recommendation is to ask all necessary and unique questions, while resisting the temptation to include unnecessary or redundant questions. Redundant questions are common in survey research when scale development follows traditional psychometric principles (Nunally [1978](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR36)), as introduced to marketing most prominently by Churchill ([1979](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR8)). More recently, Rossiter ([2002](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR42), [2011](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR43)) has questioned this practice, especially in the context of measuring concrete objects and attributes that are interpreted consistently as meaning the same by respondents. Redundant items are particularly problematic in the context of market segmentation analysis because they interfere substantially with most segment extraction algorithms’ ability to identify correct market segmentation solutions (Dolnicar et al. [2016](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR21)).

Developing a good questionnaire typically requires conducting exploratory or qualitative research. Exploratory research offers insights about people’s beliefs that survey research cannot offer. These insights can then be categorized and included in a questionnaire as a list of answer options. Such a two-stage process involving both qualitative, exploratory and quantitative survey research ensures that no critically important variables are omitted.

### 3.2 Response Options

Answer options provided to respondents in surveys determine the scale of the data available for subsequent analyses. Because many data analytic techniques are based on distance measures , not all survey response options are equally suitable for segmentation analysis.

Options allowing respondents to answer in only one of two ways, generate *binary* or *dichotomous data*. Such responses can be represented in a data set by 0s and 1s. The distance between 0 and 1 is clearly defined and, as such, poses no difficulties for subsequent segmentation analysis. Options allowing respondents to select an answer from a range of unordered categories correspond to *nominal variables* . If asked about their occupation, repondents can select only one option from a list of unordered options. Nominal variables can be transformed into binary data by introducing a binary variable for each of the answer options.

Options allowing respondents to indicate a number, such as age or nights stayed at a hotel, generate *metric data* . Metric data allow any statistical procedure to be performed (including the measurement of distance), and are therefore well suited for segmentation analysis. The most commonly used response option in survey research, however, is a limited number of ordered answer options larger than two. Respondents are asked, for example, to express – using five or seven response options – their agreement with a series of statements. This answer format generates *ordinal data*, meaning that the options are ordered. But the distance between adjacent answer options is not clearly defined. As a consequence, it is not possible to apply standard distance measures to such data, unless strong assumptions are made. Step [5](https://doi.org/10.1007/978-981-10-8818-6_5) provides a detailed discussion of suitable distance measures for each scale level.

Preferably, therefore, either metric or binary response options should be provided to respondents if those options are meaningful with respect to the question asked. Using binary or metric response options prevents subsequent complications relating to the distance measure in the process of data-driven segmentation analysis. The visual analogue scale allows respondents to indicate a position along a continuous line between two end-points, and leads to data that can be assumed to be metric. The visual analogue scale has experienced a revival with the popularity of online survey research, where it is frequently used and referred to as a slider scale . In many contexts, binary response options have been shown to outperform ordinal answer options (Dolnicar [2003](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR12); Dolnicar et al. [2011](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR18), [2012](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR19)).

### 3.3 Response Styles

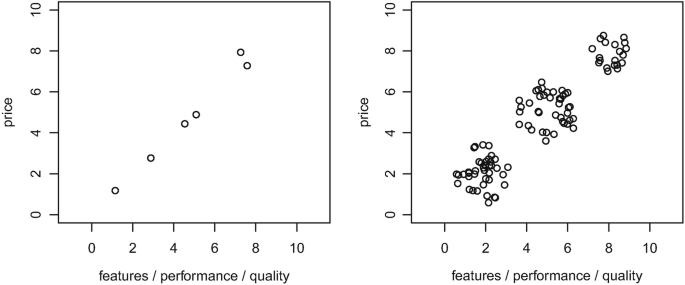
Survey data is prone to capturing biases. A response bias is a systematic tendency to respond to a range of questionnaire items on some basis other than the specific item content (i.e., what the items were designed to measure) (Paulhus [1991](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#ref-CR37), p. 17). If a bias is displayed by a respondent consistently over time, and independently of the survey questions asked, it represents a response style .

A wide range of response styles manifest in survey answers, including respondents’ tendencies to use extreme answer options (STRONGLY AGREE, STRONGLY DISAGREE), to use the midpoint (NEITHER AGREE NOR DISAGREE), and to agree with all statements. Response styles affect segmentation results because commonly used segment extraction algorithms cannot differentiate between a data entry reflecting the respondent’s belief from a data entry reflecting both a respondent’s belief and a response style. For example, some respondents displaying an acquiescence bias (a tendency to agree with all questions) could result in one market segment having much higher than average agreement with all answers. Such a segment could be misinterpreted. Imagine a market segmentation based on responses to a series of questions asking tourists to indicate whether or not they spent money on certain aspects of their vacation, including DINING OUT, VISITING THEME PARKS, USING PUBLIC TRANSPORT, etc. A market segment saying YES to all those items would, no doubt, appear to be highly attractive for a tourist destination holding the promise of the existence of a high-spending tourist segment. It could equally well just reflect a response style. It is critical, therefore, to minimise the risk of capturing response styles when data is collected for the purpose of market segmentation. In cases where attractive market segments emerge with response patterns potentially caused by a response style, additional analyses are required to exclude this possibility. Alternatively, respondents affected by such a response style must be removed before choosing to target such a market segment.

### 3.4 Sample Size

Many statistical analyses are accompanied by sample size recommendations. Not so market segmentation analysis. Figure [5.1](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5#Fig1) illustrates the problem any segmentation algorithm faces if the sample is insufficient. The market segmentation problem in this figure is extremely simple because only two segmentation variables are used. Yet, when the sample size is insufficient (left plot), it is impossible to determine which the correct number of market segments is. If the sample size is sufficient, however (right plot) it is very easy to determine the number and nature of segments in the data set.

**Fig. 5.1**

[](https://link.springer.com/chapter/10.1007/978-981-10-8818-6_5/figures/1)

Illustrating the importance of sufficient sample size in market segmentation analysis

4 Data from Internal Sources

Increasingly organizations have access to substantial amounts of internal data that can be harvested for the purpose of market segmentation analysis. Typical examples are available to grocery stores, booking data available through airline loyalty programs, and online purchase data

advantage is that such data are usually automatically generated and – if organizations are capable of storing data in a format that makes them easy to access – no extra effort is required to collect data.

The danger of using internal data is that it may be systematically biased by over-representing existing customers.

5 Data from Experimental Studies

Experimental data can result from field or laboratory experiments. For example, they can be the result of tests how people respond to certain advertisements. The response to the advertisement could then be used as a segmentation criterion. Experimental data can also result from choice experiments or conjoint analyses .

Step 4: Exploring Data

1 A First Glimpse at the Data

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

At a more technical level, data exploration helps to (1) identify the measurement levels of the variables; (2) investigate the univariate distributions of each of the variables; and (3) assess dependency structures between variables.

2 Data Cleaning

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

Returning to the Australian travel motives data set, the summary for the variables Gender and Age indicates that no data cleaning is required for these variables. The summary of the variable Income2 reveals that the categories are not sorted in order. This is a consequence of how data is read into R. R functions like read.csv() or read.table() convert columns containing information other than numbers into factors. Factors are the default format for storing categorical variables in R.

3 Descriptive Analysis

Descriptive numeric and graphic representations provide insights into the data. Statistical software packages offer a wide variety of tools for descriptive analysis. In R, we obtain a numeric summary of the data with command summary(). This command returns the range, the quartiles, and the mean for numeric variables.

Histograms visualize the distribution of numeric variables. They show how often observations within a certain value range occur. Histograms reveal if the distribution of a variable is unimodal and symmetric or skewed.

## 4 Pre-Processing

### 4.1 Categorical Variables

Two pre-processing procedures are often used for categorical variables. One is merging levels of categorical variables before further analysis, the other one is converting categorical variables to numeric ones, if it makes sense to do so.

Merging levels of categorical variables is useful if the original categories are too differentiated (too many). Thinking back to the income variables, for example, the original income variable as used in the survey.

### 4.2 Numeric Variables

The range of values of a segmentation variable affects its relative influence in distance-based methods of segment extraction. If, for example, one of the segmentation variables is binary (with values 0 or 1 indicating whether or not a tourist likes to dine out during their vacation), and a second variable indicates the expenditure in dollars per person per day (and ranges from zero to $1000), a difference in spend per person per day of one dollar is weighted equally as the difference between liking to dine out or not. 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 x¯�¯ and divides by the empirical standard deviation s.

5 Principal Components Analysis

Principal components analysis (PCA) transforms a multivariate data set containing metric variables to a new data set with variables – referred to as principal components – which are uncorrelated and ordered by importance. The first variable (principle component) contains most of the variability, the second principle component contains the second most variability, and so on.

Principal components analysis works off 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.

6 Step 4 Checklist

