

# Chapter 2

## Consumer Behaviour and Marketing Fundamentals for Business Data Analytics



Natalie Jane de Vries and Pablo Moscato

**Abstract** This chapter provides the reader with a brief introduction to the basics of marketing. The intention is to help a “non-marketer” to understand what is needed in business and consumer analytics from a marketing perspective and continue “bridging the gap” between data scientists and business thinkers. A brief introduction to the discipline of marketing is presented followed by several topics that are crucial for understanding marketing and computational applications within the field. A background of market segmentation and targeting strategies is followed by the description of typical bases for segmenting a market. Further, consumer behaviour literature and theory is discussed as well as the current trends for businesses regarding consumer behaviour.

**Keywords** Advertising · Behaviouristic segmentation · Consumer behaviour · Market segmentation · Social media marketing · Targeted marketing

### 2.1 Introduction

As Chap. 1 has already mentioned, it is time to *bridge the gap* between data science and business and marketing. As it becomes increasingly important for everyone in business to understand data analytics in more detail in order to let new technologies bring advantages, so too has it become important for data scientists to generate a level of understanding of the problem domains their technologies apply in. As Chap. 1 stated, this has to become a two-way conversation where each party to the table has the required knowledge of the other’s domain to “keep the conversation flowing”.

---

N. J. de Vries (✉) · P. Moscato  
School of Electrical Engineering and Computing, The University of Newcastle, Callaghan, NSW,  
Australia  
e-mail: [natalie.devries@newcastle.edu.au](mailto:natalie.devries@newcastle.edu.au); [Pablo.Moscato@newcastle.edu.au](mailto:Pablo.Moscato@newcastle.edu.au)

Marketing has become one of the most multi-faceted disciplines, both in academia and in the professional world and it evolves as fast as technology, trends and our surrounding world does. Although it is ever-changing, some basic fundamentals of Marketing still hold truth today and help understand the background of the field, Business and the “why” and “how” of consumer behaviour. We will cover the general background of Marketing as a discipline and slowly bring it forward to today, where business researchers and professionals increasingly rely on data analytics for success, and where data scientists increasingly find some of the most exciting applications of their technological advancements in the business world. We therefore aim to provide a brief introduction into the world of Marketing, the background marketers come from when engaging with data scientist and analytics experts in our quest to *bridging the gap*.

### ***2.1.1 Marketing as Applied, Consumer-Centric Data-Rich Economics?***

With roots in Economics, Psychology, Sociology, Science and more recently Computer Science and Mathematics, Marketing has become an extremely multi-faceted discipline. Much debate has been conducted about marketing as a discipline and the role it should play in organizations and businesses. As the marketing discipline continued to grow, a “push” was seen to bring marketing closer to the “hard sciences”, the Science, Technology, Engineering and Mathematics (STEM) subjects. Particularly in the late 1950s to early 1960s marketing literature saw a surge in research advocating “marketing as a quantitative science” discipline [69]. However, the truth is that marketing co-evolved with businesses, organizations and consumers, and has been influenced by these other disciplines. Kumar [69] specifically shows the overarching trends in marketing research and practice from the 1930s, when marketing was first spawned as a “child discipline” of economics, described by Kumar using the metaphor “Marketing as Applied Economics”, to the present day of marketing’s omnipresence in both business and life in general.

Today, marketing continues to evolve as we continuously try to understand consumers, their behaviour, their reasons and motivations for their actions and subsequently try to serve all of these consumers with goods and services that please them. What marketing is has changed over the years and cannot be answered in a single sentence. Throughout this chapter we aim to give a good overview of the area, how marketing and business units intertwine with data analytics in today’s world while also providing the necessary basics and fundamentals.

With the world-wide adoption of modern information and computing technologies and the ever increasingly interconnected world, the role of marketing in businesses is now evolving at an accelerated pace. One quote from Stephen Brown in 1996, that is still very relevant to us today and in the coming years, goes as follows:

Surely, so the argument seems to go, if marketers try hard enough, if we crunch ever-larger data sets through our ever-faster computing facilities and develop ever-more sophisticated mathematical models, we will eventually break through to the bright uplands of absolute marketing understanding [11, p. 256].

In today's data-rich world, this statement becomes all the more pressing because, how do we know we are using our "ever-more sophisticated mathematical models" or the "ever-faster computing facilities" to truly generate new knowledge and actionable insights from the hordes of data produced by everything we do and touch in our completely inter-connected world? Will we ever truly achieve "absolute marketing understanding"? Can we compare the understanding of consumer behaviour to that of the behaviour of biological systems? Probably not, but the STEM subjects have now actually been completely intertwined with business and marketing applications as part of the never-ending plight to "marketing understanding". Marketing to, and serving today's consumers successfully, requires an understanding of many aspects of scientific modelling, technological applications and mathematical foundations for data analytics.

One of the reasons for this is that consumer behaviour is becoming more and more complex with the continued adoption and penetration of new technologies, communication and purchase channels and business models [34]. Online search engines, "search engine aggregation" sites, price and product comparison sites, social networking platforms, online stores, online music and tv-subscription sites, mobile phone providers, internet providers and any company with a database store every bit of information about their consumers that they can get their hands on. Unfortunately, a lot of this information is not used to its full potential as it sits unused in endless amounts of databases or is not used in conjunction with relevant data where true benefit can be derived from it.

One consumer, in a matter of minutes, can generate gigabytes of data that are full of potentially fruitful information about that consumer. For instance, while watching a subscription online television show that same consumer is browsing an online clothing store on their laptop, skipping through songs on a recommended music playlist service, receiving instant messages on their smartphone while their wearable device keeps track of their vital health stats. Separately, the datasets that are generated from these behaviours may not provide any of the service providers with a comprehensive understanding of that consumer. However, combined with his usual browsing, listening and online shopping behaviours, the behaviours of consumers similar to him and the inter-platform data available, the information may reveal that this person is an action-movie addict, a rocker and tends to use messaging services most between certain times of day and at various intervals.

This information can be used to target that consumers' next movie recommendation, song recommendation or better targeted advertisements for the online store through his instant messaging app at the time he is actually using it (rather than untargeted ads that are irrelevant to this consumer). This in turn would cause the consumer's experience of using the online platforms and recommender sites to be more pleasurable and efficient which makes it likely for that consumer to continue to use those services. The world is now moving towards one seamless experience

between all devices and even between the “real-world” and the “online-world”. In fact, it has been found that service quality improves with personalization and a customer’s lifetime value (CLV) increases with more personalized and higher quality service [88]. We can logically imagine how much more important this is when we are connected at all times through a multitude of devices throughout our daily lives, tracking, saving and recommending almost everything we buy, do or share.

Furthermore, if targeted advertisements to this consumer using various websites are more accurate, the ads are less likely to be received as bothersome. It is for these reasons that consumer analytics has gained traction among many companies and even new business models have flourished because of this phenomenon. Companies are trying to come to grips with how to successfully derive use from the potentially very valuable information they have about their consumers and how to “monetize” their new online business models in various ways rather than the plain selling of advertisement space on their pages. Scholarly research aiming to better understand our world and the people in it is scrambling to keep up with the new technologies being generated at today’s accelerating pace. Further, staying on top of knowing how to make sense of all the new information that is generated by new technologies continues to be relevant in the online world of today and the future.

In order to deal with these large amounts and different types of data coming from many sources (also referred to as Big Data), a shift has already started to occur from having a focus on fitting models to theorized ideas, to framing theories from data-driven research [34]. A multitude of strategies are now employed by organizations and marketing researchers to understand consumers to better serve their customers and ultimately, for companies to become more profitable. Although this interconnected world looks very different now than it did 50 years ago, the basics of marketing, consumer behaviour, business-consumer relationships and strategies are still valid today. This chapter provides you with a fundamental understanding of these basic concepts covering the topics of market segmentation and targeting, consumer behaviour and specifically online consumer behaviour that is relevant to business and consumer analytics.

### ***2.1.2 The New Bridge Between Computer Science and Marketing***

Some top scientists are recognizing the key role of computer science in the twenty-first century. It all comes, hand in hand, with the increasing central role of mathematics as an ordering force aimed at “quantifying” many other disciplines. As stated, Marketing evolves as people evolve, which means the “marriage” between Computer Science and Marketing is here to stay, and to make a difference. A few years ago, Michael Mackey, President of the Society for Mathematical Biology, was quoted saying:

The conversion of biology into a more quantifiable science will continue to the extent that it might even become the main driving force behind innovation and development in mathematics... For many years the inspiration for innovation in applied mathematics has come from physics, but in my opinion, in this century it will come from the biological sciences, broadly defined [57].

In some sense, this is already a trend that we also observe between marketing and computer science, and it is perhaps much more prominent and clear. It is increasingly common to see marketing scenarios providing the source of inspiration for the development of new problems. As we have explained in the previous chapter, computer scientists *thrive* on new problems. The identification of a clearly-cut data-rich domain for which no algorithm exists that satisfies the required need attracts their attention.

The trend is not new. In the early 1990s the attention of many computer scientists was attracted as a “new type of problem” was being introduced. For these problems, decisions must be made, and in some cases resources must be allocated, even if we do not completely know what the future holds [64]. The emergence of digital libraries [91] and video-on-demand [1] were motivating this area. This means that we could interpret this situation “as if” the input of the problem is not completely known in advance and only one part of it is known at the time of making decisions. Computer scientists define *online algorithms* as the methods that they have developed for these problems. They have also defined a measure called the *competitive ratio*: the worst-case ratio between the cost of applying the online algorithm (when only part of the input is known but decisions need to be made) and the one of the *offline* algorithm (a hypothetical situation in which the entire input is known in advance and an algorithm can make optimal decisions with perfect knowledge). This led to the area of *online algorithms* and a fruitful dialogue between disciplines. Diverse applications now exist in the areas of business and consumer analytics: story scheduling in web advertising [2], display ad allocation [35], resource allocation under discounts in cloud networks [56], online pricing with impatient bidders [22], real-time optimization of personalized assortments [39] and large scale charging of electric vehicles [17] just to mention a few examples. The dialogue between computer science, marketing, business and consumer analytics is stronger than ever before and it is likely that this trend will exponentially increase with the possibilities of personalization of services and products.

In fact, returning to the second part of the quote cited above [57], it is, perhaps, a very narrow view of the world today. By the early 1980s, with the Personal Computing revolution, more and more data processing power has been coming towards the marketing, business and consumer analytics area. It is now clear that, with each of us having in our pockets what was easily considered a supercomputer in the 1980s, the time has come that these fields are going to be an inspiring muse for applied mathematics and computer science. It is probably safe to say that, when it comes to the “marriage” of Marketing and Computer Science, we are only just at the beginning!

### 2.1.3 *The Blurring Lines Between Products and Services*

To continue the dialogue of presenting common trends in business and marketing today, we talk about services. Services heavily dominate our economies. Almost everything we do online as consumers is part of us making use of a service. May it be shopping online, reading publicly available encyclopaedia information, chatting with friends or searching to book a holiday, they all mean we are using a service one way or another. Even before the widespread use of the internet, services made up a huge part of consumers' purchases (think for instance, visiting a doctor, hairdresser, going to a restaurant or a holiday). Furthermore, services have also become a lot more prominent in Business-to-Business (B2B) transactions and relationships. Companies are outsourcing business aspects such as finance, accounting, data analytics, storage and management among many others to service firms at greater levels than ever before. All of this means that services dominate a great part of the research and knowledge development in business and marketing fields. Due to the different natures of goods and services, a lot of research and theories have been conducted and developed to deal with them separately. However, the line between goods and services has continuously blurred in recent years. For example, would we call an app a good or a service? It is not physically tangible but it also is not quite a service that only is transferred from human to human. A mobile app is programmed and coded and subsequently this software is downloaded by a user. What about an online magazine or personalized training plan subscription? We cannot hold the actual software behind the app in our hands, like we can with a "typical" consumer good such as a carton of milk, however, it also does not classify as a "traditional" service such as a restaurant experience or a financial service.

In order to understand the complications of blurring lines between goods and services, we briefly cover the basic characteristics, theories and ideas about services, services marketing, how they differ from goods and how the marketing of each typically differs.

#### 2.1.3.1 **Services Marketing Fundamentals**

We traditionally split products into two categories: goods and services. That means, physical products we can hold, and services provided to us that add a benefit (or *value*) to our lives. The difference between the two was described by four key attributes: tangibility, inseparability, heterogeneity and perishability [74]. In a pre-internet age, these four principles made a lot of sense and applied to almost every service. For example, being served in a restaurant is not a tangible object you can hold, for the service is *inseparable* from the moment you are consuming it and inseparable from the service provider. An experience in a restaurant is *heterogeneous*, as you may have a different experience every time you go into a restaurant, depending on the day, the different person serving you, etc. Finally, it is *perishable* as you cannot "save" it and consume the experience at a later time.

However, as goods and services have become increasingly intertwined, these four principles may not always be true nowadays. Today, services are “heterogeneous” on purpose. Meaning, they are personalized to your benefit, made to suit you and to generate as much *value* as possible for you personally. Further, some of them are not really perishable. For example, your mobile smartphone, your smart television device or simply your computer can remember exactly where you were up to in your latest Netflix “binge” or music tracking playlist.

In response to this, more recently, Lovelock and Gummesson have introduced a different view of services which posits that “services offer benefits through access or temporary possession, instead of ownership, with payments taking the form of rentals or access fees” [73]. This type of view is more generally applicable in today’s online world. For example, temporary online movie rentals, book rentals through e-Reader devices, online music and tv-show streaming services all fall under this way of viewing “services”.

Although the lines between “products” and “services” have blurred, there are still several characteristics of services that make them different from products that are relevant today, some of them highlighted by Lovelock and Gummesson in a more recent work [74]. Services cannot be inventoried. This has an impact on businesses and strategic marketing managers for forecasting demand. In terms of the online economy, this means, for example, enough computing power needs to be made available for enough customers to be able to use the software at once. For instance, if a website crashes when too many people are trying to use it one time, the company will likely have many dissatisfied customers. Likewise, if a company has already made surplus computing power or data storage space available for what they thought would be a high-demand time, those resources could be wasted resulting in financial losses for the firm.

The aspects that cannot be touched, seen or visualized are usually what creates value when it comes to services. These aspects are what impacts on consumers’ decision making processes, a company’s product and service development teams or the way it is marketed. This is one reason why customer online reviews, and other forms of Customer-to-Customer (C2C) interaction (see, e.g., Sect. 2.3.1), have exploded in popularity in the past decade. Reviews and consumer’s experience statements are nothing new in marketing and advertising. However, the level of online reviews available to consumers today is unprecedented. We all know *TripAdvisor*<sup>1</sup> or similar customer review websites reviewing anything ranging from travel services and restaurants, to hairdressers, transportation companies, video games, or even doctors and other professionals’ services. It is these kinds of online behaviours that have now become part of the customer service experience in many industries and for endless amounts of services (both online and offline).

---

<sup>1</sup><https://www.tripadvisor.com>.

## 2.2 From Direct Marketing to Market Segmentation and Targeting Strategies and Back Again

Marketing and consumer behaviour textbooks have typically taught us that market segmentation and targeted marketing strategies are necessary for a successful relationship with your consumers. They have also taught us that “direct marketing” is usually too expensive and not usually a viable alternative. However, with almost every consumer in developed countries having access to the internet in the palm of their hands or on their wrist (or in the near future, maybe even under our skin!) and the capability of communicating with businesses and fellow consumers, this has changed. In the mobile and online market place, direct marketing and advertising is very real, is happening in real-time and is saving consumers a lot of time while making businesses more money.

However, before we discuss today’s situation, we take a brief history lesson about segmentation in the marketing discipline and how and when market segmentation came about. Targeted marketing strategies are explained and their benefits to businesses and how targeting to customers is done now by online companies and brands. After, we return to today’s setting of direct marketing capabilities, unprecedented levels of personalization and even prediction.

### 2.2.1 *A Little History to Market Segmentation*

Market segmentation is the process of the market being divided into smaller groups of consumers with similar wants, needs and characteristics and who may require different goods and services or alternative marketing and advertising strategies [40]. Market segmentation emerged in the literature and in business in the 1950s (see, for instance [26, 93] and [33]). Marketing as its own discipline started to gain more traction and marketers wanted to gain better insights into their customers. In fact, Smith [93] was one of the first to put market segmentation forward (in 1956) and it has since been defined in many different ways.

Although it may not previously have had a name or label, market segmentation has existed for as long as businesses and suppliers have used different methods to gain clients. However, in the last five decades, marketers have been able to segment better, thanks to the development of new economic theories and sophisticated analytical methods [28]. In the early 1960s, business and marketing literature began referring to market segmentation as a new actual business “strategy” as it gained more popularity and attention among scholars and professional marketers [10, 38, 76, 87]. Companies had realized that mass marketing (think: Ford’s Model T in the 1920s) was not as successful as when Henry Ford first started selling his black Model T’s through mass production. Marketers began to understand that each individual consumer is different and expects to be served according to their own needs. Scholars started analysing, discussing and explaining various approaches to segment a market. People have different purchase behaviours because they have



different tastes, different ways of expressing themselves and varying expectations from businesses and brands. A successful marketer has to incorporate all of this into their market segmentation strategy. As Roberts [87] states, market segmentation could be described as a strategy with the philosophy of “something for everybody”.

However, it is exactly this “something for everybody” philosophy that is hard to achieve in real applicable market segmentation. Finding out how to define market segments is an aspect of market segmentation with the greatest applicable consequences [5]. The many variables to choose from make market segmentation no easy task [41]. The purpose of market segmentation is to target the segments that are found individually with an altered or specialized marketing and advertising strategy. Therefore, the way we segment is going to decide what promotional material and communication the consumers are going to receive from the company or even the next line of products a company may develop and manufacture or the kinds of services it will add to its offerings [60]. An extensive work on Market Segmentation can be found in Wedel and Kamakura [104] however, for an understanding of the market segmentation bases, we will cover in summary the most well-known bases of market segmentation.

## ***2.2.2 Different Ways of Cutting the Consumer Pie***

We cannot market the same thing in the same way and at the same time to all consumers. This is why the consumer market needs to be split up (segmented) in order for marketers to target each segment. To describe the different bases of segmentation here, we have taken the broad categories outlined previously by Schiffman et al. [89]. In turn, these segmentation approaches are highly similar to earlier works such as Beane and Ennis’ market segmentation review in the late 1980s [7] and many other market segmentation works. Both of these works include the following bases for segmentation: geographic, demographic, psychological, psychographic, sociocultural, user-related, user-situation and benefit segmentation as well as possible hybrid segmentation methods.

### **2.2.2.1 Geographic Segmentation**

The key idea behind geographic segmentation speaks for itself. It has been happening for as long as people have exchanged goods and services. When starting any sort of transaction, for many years, the only possibility has been to exchange locally or within reasonable travel time. Before the advent of the internet and e-commerce, businesses had no choice but to geographically segment their market. Although this has changed in today’s online society, geographic segmentation still exists and is still practised by businesses in conjunction with other segmentation criteria. When segmenting geographically, factors such as housing density, climate (weather) and relating factors still need to be taken into account [89].

### 2.2.2.2 Demographic Segmentation

As the name suggests, demographic segmentation is a partition of the market based on a consumer's specific demographic information, such as their age, sex, marital status, income, education level and occupation [89]. This is also one way of segmenting that is age-old and has happened even before scholars coined the term "market segmentation". Any product that is specifically for males or females or for children or adults has used the concept of demographic market segmentation. Furthermore, a lot of research is done on the behavioural and spending patterns of consumers with different levels of incomes.

### 2.2.2.3 Psychological Segmentation

From the surface, psychological segmentation is difficult to define. It refers to the inner or psychological characteristics of consumers, their intrinsic qualities [89]. Endless amounts of qualities could be associated to any given person which makes it extremely difficult to drill down to several qualities that are able to segment the complete market. Some of these qualities include consumers' motivations, personality, perceptions, learning, level of involvement or their attitudes. As market segmentation theory continued to evolve, marketers wanted to "dig deeper" into the minds of their consumers. Along this line, psychological segmentation surfaced as a way to define, describe and segment the market in a way that geographic or demographic variables could not.

This "new" and exciting alternative approach to segment the market was promised to "shake up" market segmentation was traditionally conducted. However, although psychological constructs are "rich" in information, they are often unreliable when investigating large groups of people [85]. In order to still use psychologically-related variables, marketers have come with "life-style analysis" which has proven to be a useful marketing tool. Lifestyle analysis is part of psychographic segmentation [89].

### 2.2.2.4 Psychographic Segmentation

Psychographic segmentation is also referred to as *lifestyle segmentation*. The concept of consumer lifestyle patterns was first introduced to marketing by Lazer in 1963 and in the later 1960s and the early 1970s became more popularly used as a basis for market segmentation [85]. As Plummer [85] states, some of the most commonly used lifestyle segmentation approaches back then were activities, interests and opinions. Another argument for psychological segmentation is that it may be more generalizable than demographic or geographic segmentation as the segments based on psychological characteristics could exist in multiple markets [70]. Although it seems like a logical way to segment the market, it has been difficult to apply these variables to large groups of people. However, nowadays with

the widespread use of the internet by almost all customers and business, gathering information about these aspects has become more feasible. More on this later.

### **2.2.2.5 Sociocultural Segmentation**

Sociocultural segmentation, as its name suggests, takes a cultural view of segmenting the market. Examples of such characteristics include stage of family life cycle, social class, core cultural values, subcultural memberships and cross-cultural affiliations [89]. As the family life cycle progresses, it is natural that the household will require different products and services (think single person to married couple with kids, from babies to toddlers to teenagers and so forth).

The other part of sociocultural segmentation is the cultural characterization of a consumer. Consumers' values and attitudes have been found to have an effect on their purchase behaviour and other consumer behaviours towards brands. As Vinson et al. [101] explained as early as in 1977, a person's attitudes and values are affected by their sociocultural surroundings. It is for this reason that cultural characteristics make an interesting segmentation base that provides rich information for segmentation purposes. However, again, its downfall is that it is difficult to apply to large groups of people due to its lacking generalizability and obvious concerns for possible discrimination occurring when segmenting based on social status or cultural backgrounds.

### **2.2.2.6 User-Related Segmentation**

Rate-of-usage segmentation takes into account how much a consumer purchases or uses a product or service and a segmentation is made accordingly [89]. It has also been referred to as usage-rate segmentation [7]. Beane and Ennis [7] explain that usage rate segmentation may divide consumers into light-, medium- and heavy user groups. Marketers then generate separate marketing and targeting campaigns for each. One common example of this is insurance companies who give discounts for "less usage". This gives customers the feeling that the services are more personalized to them and that they have more control over the products they choose. This type of segmentation, where applicable, has now become extremely common and is intertwined with many other segmentation and targeting strategies.

### **2.2.2.7 Benefit Segmentation**

Benefit segmentation can also be called "needs-based segmentation" as it refers to the actual need the product or service satisfies. It is also part of "behaviouristic segmentation" as covered by Beane and Ennis [7] and earlier by Kotler [67]. Companies have been using it since 1961; however, research about benefit segmentation lagged behind. Haley [50] describes it as an alternative, and more optimal base for

segmenting the market. As he states, the belief that underlies this strategy is that the reason *why* a consumer chooses a product (or the benefits that they seek) are the true underlying reasons for the existence of market segments.

With benefit segmentation, the idea is to look at the segments that surface after consumers are divided into groups of similar benefits and needs followed by an examination of their demographic information, volume of purchase and other aspects. In doing so, a richer image can be created of the consumers. This strategy is still relevant today. For any goods or service provider it is important to understand the benefits their consumers are seeking from their products and the needs they want to have satisfied. When these questions are answered, part of the question of *why* that consumer has chosen to purchase that particular good or service is known.

### **2.2.2.8 Hybrid Segmentation**

The last segmentation basis that Schiffman et al. [89] discuss is that of hybrid segmentation. In Beane and Ennis [7] several mixtures of segmentation approaches are also discussed. For example, “componential segmentation”, which combines situational variables with consumer/respondent characteristics. Hybrid segmentation in general can be a mixture of some or all of the above-mentioned segmentation bases. This type of strategy is closest to what companies have done in real-life by combining several bases for segmentation to segment the market more accurately and target more efficiently.

## **2.2.3 Targeted Marketing and Advertising Campaigns**

As stated, the main driver for segmenting the market is then to target each segment with a varying strategy. With the continued technological developments available, businesses have been able to implement narrower targeting strategies to target various consumers segments more precisely. Generally, even after the market is segmented, a lot of communication material or promotional content will reach people that it is not intended for, or who are not interested in it. For example, television or radio advertisements where one ad is shown to a large group of segmented people (for example, a cooking appliance ad during a cooking show), could reach a large section of that group may not be interested in purchasing a cooking appliance any time soon. With the increased online information available about consumers, targeted marketing and advertising campaigns can be made more narrow and precise resulting in more successful marketing strategies.

Furthermore, scholars and marketers now understand that we cannot treat all consumers in even the same segment the same. Consumers that have been segmented based on certain geographic, demographic or even psychographic criteria may have vastly different customer behaviours towards brands. High levels of heterogeneity even exist when examining consumers' online behaviours towards brands within

customers of the same brand [23]. This shows to brands that, although they have a certain brand identity that they try to target consumers with, who themselves identify with that “brand personality”, they cannot treat all these consumers in the same way. This means that we need to look at consumer behaviours (See Sect. 2.3 on Consumer Behaviour) when segmenting and targeting the consumer market and go beyond the “traditional” segmentation bases and even beyond known targeting strategies. This is where modern-day personalization comes into play.

## 2.2.4 “Back Again”: Direct Marketing and Personalization

Personalization to and direct marketing at the individual level are now a reality. Although market segmentation and targeting strategies are not completely obsolete, they are now complemented by highly effective and detailed personalization analytics and artificial intelligence methods.

### 2.2.4.1 Direct Marketing Made Possible Through Technology

In some sense the direction for personalization of offers has made “direct marketing” more viable. However, this is not a new concept. Direct marketing is a form of advertising that has a strong focus on the customer and employs data and testing strategies. Its defining characteristics are that: (a) the messages are directly addressed to customers and/or prospective customers, (b) there is a “call for action” (it used to be a call free phone number or mail order, now it could be a click on a website) and; (c) there is an emphasis in quantifiable/measurable responses and in monitoring response rates (and objectively tries to minimize future advertising expenses while maximizing the response rates). It is a strong data-driven procedure in which pre- and post-campaign data analytic procedures are used.

This “learning-from-data” approach is certainly employed by all types of businesses nowadays. Companies like Amway, Avon, Herbalife, Markay, Vorwerk and Natura are leaders in total sales (in US dollars) using this approach. The total revenue of the top 100 companies is approximately 81 Billion US dollars in 2015.<sup>2</sup> Furthermore, direct marketing has also been central to the activities of many charities and non-for-profit organizations as a strategy to address people directly with compassionate messages [16].

It is reasonable to assume that the data analytics methods that were normally employed in direct marketing were heavily guided by “simple” statistics. Interventions in the marketing campaign were designed and the outcome evaluated *after* the fact. With the advancement of new technologies and computer science methods, in particular machine learning, we are now observing some changes. Quantitative

---

<sup>2</sup><https://www.businessforhome.org/2015/04/top-100-global-direct-selling-companies-2015/>.

direct marketing models are now statistical and/or machine learning based [9]. Heuristic and metaheuristic techniques are being employed due to the large size of the datasets [82]. Marketers and advertisers now aim to take a more predictive approach in their strategies. This is possible due to the large sizes of datasets marketers now have access to.

Besides large datasets, what brings real value to marketers, researchers and business decision makers is the high variety and velocity of datasets and the combination of these sources of available information. Consumers' purchase and personal data has become extremely valuable and opinions of its use are highly contrasting. The next section covers the personal data use trend and aspects.

#### **2.2.4.2 Personalization and Monetization in the Personal Data Business**

By now, most consumers in the developed world will have used some form of product or service that is the result of today's great technological advancements making high levels of personalization possible. Think for example, Netflix, Spotify, an online booking system and so on. All of these services run on algorithms that include recommender systems, clustering, prediction and more in order to personalize the service to your experience. Furthermore, even if you have not subscribed to any of these services, but may have simply searched the internet, used a social media networking site or bought something online as a once-off customer, you have likely still been a consumer of a personalization strategy. Furthermore, as stated above, it is almost certain that the personal information that you have provided in these online activities and transactions has been recorded by one or multiple companies and stored for future marketing or advertising purposes.

#### **2.2.4.3 Personal Data: Online Advertising**

In recent years, data and (consumer) information has popularly been referred to as the "oil of the twenty-first century" (and some say this analogy has become a bit of a cliché). Having and being able to store, analyse and make use of extremely large amounts of business and personal data is now seen as extremely valuable. Not only for the company from which the data came from, but for many third-party companies or advertising companies.

One of the main and first reasons for this current personal data use and monetization "wave" is online advertising. Whole business models are now based on the notion of making profits from advertisements. Online advertisements are at the forefront of aiming to predict a consumer's next choice in order to make financial gains through their pay-per-click strategies, for example. Naturally, the more information available, the better the advertisements can be predicted and the more successful they are likely to be. One of the world's most popular social and digital companies, *Facebook*, has as its main "business" advertising. In fact, 2014, 2013 and 2012 advertising accounted for 92%, 89% and 84%, respectively,

of Facebook's revenues [3]. Furthermore, Google in 2014 reported \$66b in total earnings, of which \$59.62b came from advertising alone.<sup>3</sup> Since 2014 these figures are likely to only have grown and they provide a good indication of the size and value of the online advertising industry.

#### **2.2.4.4 Personal Data: Data Management**

Advertising is not the only reason why companies collect and store our online (and offline) movements. Other ways in which personal data is used and monetized are for marketing and product development activities. For example, knowing how customers are actually using your online store, service or application can provide managers with useful insights that can feed into the product design stages and general marketing strategies of the company. If we take a popular social networking company like Instagram. Instagram actually started out as a company for consumers to “check-in” locations, share plans, information as well as photos. That is, photo sharing was only one feature of the first app. However, after learning how consumers were actually using their app, it became apparent that the photo-sharing feature was the most popular even though location-based applications were “all the rage” at this time.<sup>4</sup> Subsequently, they re-branded, changed their actual service and became what Instagram is today. Without the consumer insights coming from their usage and behavioural data, this would not have been possible.

More recently, extremely large amounts of data have become valuable for data management purposes. With the large deluge of data that companies and organizations find themselves with, having the capability and know-how of how to store and manage these amounts of data safely also becomes a valuable asset. Cloud computing companies' business model is to generate profit from providing storage space and computing power to businesses on an on-demand basis. In recent years, these types of services have become much more prevalent as the demand for them is ever-increasing. For many companies it simply is not feasible to store and manage the large amounts of data themselves as they do not have the resources available. This is where the value lies for storing and managing big data.

#### **2.2.4.5 Personal Data: Ethics and Privacy**

As expected, huge debates regarding the ethical implications of companies using, storing, analysing and selling consumers' personal data are happening everywhere. Consumers are increasingly more concerned with the safe storage and different uses of their personal data [20]. In the personal data debate we have two sides. On one

---

<sup>3</sup><https://investor.google.com/financial/tables.html>.

<sup>4</sup><http://www.investopedia.com/articles/investing/102615/story-instagram-rise-1-photo0sharing-app.asp>.

side are consumers who want to protect their privacy while being able to make use of the latest and greatest technologies, while on the other side we have companies who want to be able to make a large as possible profit from providing these technologies and doing so requires the use of their consumers' personal data.

Further, as the Accenture report explains [20], privacy is an inherently personal concept and a "one-size-fits-all" approach is not going to work in the prospect of protecting privacy in today's changing technological environment.

### **The Ethics of Personal Data: A Hot Debate**

A somewhat recent example that was the topic of a "hot debate" about the use of personal data comes from the messaging "app" *Whatsapp*. After being bought by Facebook, Facebook's promise to consumers was that conversations within the Whatsapp "app" would remain private and would not be used for Facebook's personalized advertising campaigns. However, in August 2016, Whatsapp<sup>5</sup> announced a change in terms and conditions when users downloaded a new update which in fact would share personal information with Facebook unless consumers opted out before a specific month. Many users were angry and frustrated as Facebook Executives had previously proclaimed that "nothing will change" and that Whatsapp will remain independent when Facebook bought out the popular messaging service in 2014.<sup>6</sup> Many online bloggers and users of the app expressed their frustrations online meaning Whatsapp's PR team probably had some very hard times. This example shows how companies must take making promises to its users about the use of personal data very serious. We are yet to see the end of the story as the legal ramifications of these decisions continue.<sup>7</sup>

From a legal point of view, it is also important to see some progress in terms of how online communications, transactions and transfers of information and data are viewed. As Hasty [51] explains, digital interactions could (should) be framed as commercial exchanges of value in the eyes of the law. Many ethical and legal changes are going to be necessary with the ever-changing trends in the online personalization, advertising and information sharing environment. One main trend that we are seeing now is that consumers are increasingly seeking a data dividend as they are realizing the potential monetary value of their own data. Some call this the "second wave".

<sup>5</sup><https://www.whatsapp.com/>.

<sup>6</sup><https://techcrunch.com/2014/02/19/facebook-buying-whatsapp-for-16b-in-cash-and-stock-plus-3b-in-rsus/>.

<sup>7</sup><https://www.reuters.com/article/us-eu-dataprotection-whatsapp/facebook-eu-regulator-says-whatsapp-yet-to-resolve-data-sharing-issue-idUSKCN1GC2H8>.



### 2.2.4.6 Personal Data: “The Second Wave”

We are now on the cusp of the monetization trend changing to its second “wave”. Consumers and businesses are each coming up with new alternatives to solve the “privacy debate” changing the course of personal data monetization. Besides the practical and economic motivations, the notion of giving consumers back the power over their own data is receiving increasing attention from various communities due to the many ethical and privacy-related issues with using personal data. As Peter Sondergaard, the Senior Vice President at Gartner Research explains, in this second wave, consumers will be enabled and empowered to own and thereby monetize their own data. Meaning that they can take back the control and drive up value for themselves.<sup>8</sup> Sondergaard predicted this trend in 2013, however now, this is actually happening.

In a recent report, Accenture found that 60% of their survey respondents reported to having engaged in activities to monetize their own data [20] showing that this is a growing trend. In return for this growing demand, companies are starting up providing exactly this service to consumers. For instance, *People.io* is a platform that allows consumers to sign up and benefit from “giving away” their personal data. The founders of *People.io* explain that one of the main motivations for this platform, and one of the main reasons for its (potential) success is the extremely high increase in use of ad blockers.<sup>9</sup> It is important to highlight the main difference between a consumer signing up through a platform like this, and simply “giving away” data to the companies they are already purchasing from or receiving discounts in return for information. The difference is that this platform will *license* the consumers’ data, time and time again. A once-off provision of your personal information to a web-based company may only prove to be worth 50 cents (maybe even less depending how valuable of an eCommerce customer they deem you to be). However, the value lies in this information being used time after time aggregated with information from more sources and combined with information of other consumers. This is how online personalization advertising companies currently generate a profit. It is for this reason that letting consumers “licence” their own data out in a similar fashion, it would actually prove beneficial for those consumers.

The dialogue (and the online power play) will now continue to find an interaction between companies and consumers that will aim at creating a mutually beneficial outcome for all. One where consumers’ do not feel like their privacy is intruded, but one where each individual determines their level of data sharing or “licensing” and where personalization does not equal privacy “invasion”. Daniel Newman states the following in a Forbes Magazine article which, perhaps, states this next stage (possibly the “third wave”?) of online personalization:

---

<sup>8</sup><http://blogs.gartner.com/peter-sondergaard/monetizing-personal-big-data-second-stage-of-digitalization/>.

<sup>9</sup><http://econsultancy.com/blog/67441-start-me-up-people-io-allows-people-to-monetize-their-personal-data/>.

Personalized, data-driven marketing will become more refined. There is a difference between data-driven marketing and intrusive marketing. While the former is based on relationship-building, the latter is nothing but old-school push marketing wrapped in a new cover. The difference between these two formats will become even more prominent in future. Marketers who focus on relationship building will be rewarded, while intruders will be shut out.<sup>10</sup>

## 2.3 Consumer Behaviour and Interaction

Consumers' varied and heterogeneous behaviour is one of the main drivers for market segmentation, targeting campaigns, direct marketing and personalization strategies [25]. Consumer behaviour is defined as the behaviour that consumers display (towards a brand or business) in searching for, purchasing, using, evaluating and even disposing of the products and services that they expect will satisfy their needs [89]. The key idea to take from this definition is that consumer behaviour does not necessarily mean purchase behaviour. It can be any display towards a brand in the form of an online comment or "like", sharing something to friends, blogging about a certain product or service, rating and reviewing services online and so forth. The current online nature of interactions between consumers and businesses has meant that it is possible to track and analyse a lot more non-purchase behaviour of consumers. At the same time, it has also amplified the effects that non-purchase behaviour can have on a company's success (both positive and negative). When we talk about consumer behaviour prediction, online recommendations and personalized advertisements, it all centres on being able to understand consumer behaviours. This is why we bring back some marketing fundamentals and present the broad topic of consumer behaviour and interaction.

### 2.3.1 *Changing the Way We Communicate with Consumers*

Customers can *engage* with brands online and display many different behaviours towards their brands of choice without ever having to purchase anything [24, 100]. Historically, businesses and organizations only communicated to consumers and consumers were always the receivers of communication messages, advertisements or promotional material. With the continued development of the marketing discipline among scholars and the continued acceptance by businesses of marketing as a "true" business unit, a trend was heralded by prominent scholars such as Kotler who continued some of the earlier views that Drucker held about the role of marketing in businesses and economies [31]. Marketing started moving away from

---

<sup>10</sup><http://www.forbes.com/sites/danielnewman/2015/04/14/10-top-trends-driving-the-future-of-marketing/2/>.

a heavy emphasis on price and distribution and move to a greater focus on meeting customers' needs and on the benefits received from a product or service. Kotler [67] was one of the first (in 1965) to recognize that marketers needed to “dig deeper” to generate a better understanding of their consumers. All of the more advanced segmentation, targeting and personalization methods that we have discussed would not be possible without a sound understanding of consumer behaviour. This trend continues to evolve as the society and economy we live and operate in continues to be “disrupted” by newer, faster, better and more all-encompassing technologies.

While Drucker and Kotler were looking at purchase behaviour of consumers at physical stores and service outlets, we now deal with consumers who can be interconnected 24/7 and are able to make purchases at any time that suits them. There are many more possibilities of what constitutes as *consumer behaviour* towards brands than in Drucker and Kotler's time. Whereas previously, the only real behaviour marketers could analyse was purchase behaviour. Now, there could be years of information on and interaction with a consumer, without that consumer in fact having purchased anything from that brand. A popular joke these days on the Internet goes like this:

A million guys walk into a Silicon Valley bar. No one buys anything. Bar declared massive success.

Another example of non-purchase consumer behaviour is a motor enthusiast who may follow everything a particular auto brand does and interacts with them online but cannot afford to purchase an old-timer classic model themselves. They can follow, like, comment, interact with and share any information of and with this brand, be an advocate for this brand and have a significant impact on the brand's online communication success without being a customer themselves. This “online playing field” brings a whole new area of both challenges and opportunities to marketers. Behavioural data has become a gold mine of information for those users that know how to extract it. We now need to see consumers as partners in a lasting relationship, and serve them accordingly. This book will feature many different aspects of analysing the purchase cycle of consumers or the business product development cycle and its related behaviours. Further, we now do not only look at communications and behaviours between consumers and companies, but also among consumers. Customer-to-customer (C2C) interactions have become a popular area of research for marketing and business practitioners and researchers.

### ***2.3.2 Customer-to-Customer Interactions***

It is obvious that with the widespread adoption of new technologies such as the internet, smartphones, tablets, wearable internet-connected devices and others, communication has changed. Typical business-to-consumer (B2C) communication still happens very frequently, as well as business-to-business (B2B) communication; however, consumer-to-business (C2B) and consumer-to-consumer (C2C) commu-

nication have become just as common. Whereas in the past, a dissatisfied (or an extremely delighted) customer had to go through the effort of returning to the store, service outlet or organization location, or send a letter via the post to express how they feel, nowadays, leaving a review, a recommendation, complaint, compliment or a “thumbs up” or “thumbs down” are only a few clicks (or swipes) of effort away and can easily be seen by hundreds or thousands of other potential customers. Specifically, the C2C form of communication has potential to provide businesses with actionable insights if used effectively. Further, C2C interactions can also be represented by networks when these consumers are connected in some form, for instance, a Twitter follower network, or a Facebook friendship network. As you will find in this book, a network can be investigated, explored and analysed in a multitude of ways to provide meaningful insights to business decision makers and marketing managers. Specifically, Part III focusses on network analytics applications. As part of looking at C2C interactions, user-generated content and word-of-mouth communications are highly important.

### **2.3.2.1 User-Generated Content and Word-of-Mouth**

C2C interactions and communication are extremely important to the success of a business as they can affect growth and profitability of a business as well as its reputation [72]. This also means that marketers and researchers now have access to more information about consumers than ever before. Online review platforms, discussion and help forums and online blogging platforms are rich sources of information where consumers often “speak their mind”. Consumers in fact generate a lot of content whether it is through reviewing service providers such as restaurants or by simply uploading photos of their experience. These types of behaviours all become part of the C2C interactions that could affect the success of a business depending on the nature of the content being uploaded or the consumers’ experience which they share with fellow internet users.

Consumers are more likely to trust communication about a brand coming from a fellow consumer than from the brand or business itself. Although WoM research is nothing new, the speed, levels of interactivity and times at which C2C interactions now happen are. Most review or booking sites (for example, Tripadvisor) now provide mobile application versions of their service (website) making it even easier for consumers to leave a quick review. Consumers are now also more likely than ever before to actually generate content in the form of photos, videos or blog posts about the service provider they recently used. This provides the business with both genuine content that they can use in their own communication strategies and a real insight into consumers’ minds and their opinion of that business or brand. With some clever social media strategies, service providers can “ride” of the back of user-generated content and WoM without having to employ many blog writers themselves or hiring a team of photographers.

When a social media strategy that encourages consumers to take action themselves becomes extremely popular and successful, it has been termed the somewhat cliché “viral marketing” strategy.

### Using “Viral Marketing” for Social Good

Online “viral” campaigns can also be used for social good. A great social movement like the “Ice Bucket Challenge” in 2014 would not have been able to occur without the widespread use of social media among so many different demographic groups. Although “The Ice Bucket Challenge” received some negative press about creating a “hype” rather than *actually* helping people with ALS, the movement proved hugely beneficial for ALS charities and their fundraising. According to the (US) national chapter of the ALS Association (ALSA), the challenge brought in a staggering \$115 million. Participants also donated an additional \$13 million to the association’s regional branches. Normally, these kinds of numbers were unheard of for the ALSA. The charity’s official form filings for 2013 show they brought in \$23.5 million that year, meaning that in 2014, there was a national donation increase of nearly 490%! Since then, ALS researchers have actually been able to make a scientific breakthrough as a direct result of research that was made possible due to the large amount of money raised.<sup>11</sup> If that does not prove that successful viral marketing campaigns can make a “real” difference, no campaign will.

As campaigns such as the Ice Bucket Challenge have shown us, consumers can now engage with brands online through any number of technological portals, at anytime and from anywhere [24] at a never before-seen ease of access. In all of these online communications, “likes”, “follows”, “retweets”, “favourites” and so forth show how many options consumers have to interact with each other and with organizations.

The greater use of, and dependence on, social media in communication also means that consumers themselves have become more demanding. When a consumer poses a question or complaint to a brand’s Facebook page, they expect a swift response as this is the nature of online social communications. Gone are the days where you had to wait several days for an email in return or even a slow response on a review website. If you are unhappy with a company’s service, their product, or how long you have been waiting in their telephone queue, you simply find them on social media, “slander” them publicly and they will automatically feel compelled to

---

<sup>11</sup> <https://www.theguardian.com/science/2016/jul/27/how-the-ice-bucket-challenge-led-to-als-research-breakthrough>.

respond in a timely manner in order to save face. However, due to this quick and easy online nature of C2B and C2C communications, brands can also pleasantly “surprise” their consumers by taking quick action.

Pleasing, or “delighting” customers is all about “managing expectations”. Any first-year (services) marketing textbook will tell you that if a company fails to deliver the customer’s basic expectations, that customer will be dissatisfied (see, for instance, [74]). Oppositely, if a company or service manages to exceed the customer’s expectations, he/she will be delighted. This same theory applies to social media communication. Social media platforms themselves can be seen as a service, or at least a service platform [24]. By this we mean that the basic theories of service marketing also apply when communicating with consumers online. Therefore, if a customer is particularly unhappy and expresses so in an online post or comment, the company now has an opportunity to revert this situation into a positive one. It is likely the customer is not “expecting miracles” from their online communication which means that if the company is able to resolve the issue, the customer’s expectations were exceeded, making them a delighted customer.

It is extremely important for a brand or company to understand their customers (and potential customers), so that they can “delight” them in the time that these consumers spend on the online platforms of the brand. Through today’s advanced analytics capabilities businesses have become more successful in understanding their customers, or target market, and personalize their offerings accordingly.

## 2.4 Empirical Research in Marketing

As the whole theme of the first two chapters of this book is to *bridge the gap* between Marketing and Computer Science, we cannot leave out a section on the typical existing empirical research conducted by marketing and business researchers. In our search of ever greater “marketing understanding” as Brown has stated in the quote at the start of this chapter, marketers and researchers are moving closer and closer to the exact sciences and use ever increasingly complex mathematics to continually improve market understanding and knowledge. It is due to this that the Marketing discipline so heavily adopted Statistics in its basic workings. It is here that we will find how much Marketing has been influenced by the discipline of Psychology which in turn has had a long-lasting “marriage” with Statistics. It is thanks to statistics that researchers have been able to understand human and consumer behaviours in more depth through methodological analysis of data. Marketing has been influenced and borrowed methods from other disciplines, more recently psychometrics and mathematics [15] and computer science coming in the later years to make a difference.

If we look at the last few decades of economic trends and human developments, we have seen an immense improvement of our understanding of the world around us. Not just the physical world, but also the more nuanced world that is human behaviour and human personalities. In these trends, we have become obsessed with

knowing, understanding and even predicting the world around us. From weather forecasts to financial forecasts to predicting whether Amy browsing on her mobile device is going to make the purchase in her shopping basket. As Marketing is a fairly young discipline (compared to its STEM relatives and even Psychology), it has only been around since the human drive for more and greater knowledge has grown to unprecedented levels. This means that marketing and business researchers have aimed at applying the latest and greatest research methodologies and the best statistics to generate a “greater marketing understanding”.

As the internet has only been used by a widespread amount of people for the last few decades, prior to this, researchers had to collect data in different, more traditional, manners. Whereas now, companies have gigabytes worth of data on their customers, previously this was not the case. So, if a Telecom company, for instance, wanted to understand its customers intentions and customer satisfaction better, the best way to find out was through survey research. The researchers would put together a carefully selected questionnaire that would comprise of *constructs* that would aim to get a good grasp of the theoretical concepts such as customer satisfaction. Each construct may be made up of a number of variables that do the best job of truly collecting information that reflects the theoretical concept. Surveys would be collected from a large enough number of customers and this data can then be used for statistical analysis to create useful business insights. Hence, most marketing research that you will see conducted in recent decades will be survey-based research.

### ***2.4.1 Using Surveys in Marketing Research***

Survey research is most definitely not just a thing of the past. Even though we have a deluge of data available to us these days, sometimes, the specific questions we want to ask cannot be answered with existing available data. This is where survey research still holds an important place in today’s market research space. Direct, targeted and specific information can be obtained from consumers (with their given consent) easily, quickly and at a low cost. They also still provide a good starting point for companies to find out initial insights about customers or, on the other hand, in order to dig a little deeper into an issue that a company might be facing (e.g. why are customers leaving the company?). Due to the still many reasons why survey research is still relevant today, we will cover the basics that go into empirically setting up survey research projects. A good study for the interested reader is that by Malhotra and Grover [75] on survey research. They explain that a survey typically has three characteristics: *asking people* for information in an organized format, a *quantitative method* and a *sample* (i.e. only a fraction of the population of interested is investigated) [75]. These characteristics bring with them important considerations. Asking people information in a structured approach can be done in many different ways (offline, face-to-face, email, etc.), the quantitative method used needs to be the appropriate one for the study and for the specific

research questions that need to be answered, and finally, it needs to be possible for the findings based upon information from the sample to be generalized for the whole population. In explaining these characteristics and the important considerations for survey research, what we are really saying is that we want to avoid GIGO; as Churchill [19] stated, marketers need to avoid spending too much effort on what computer specialists sometimes call “Garbage-In-Garbage-Out”. Hence, following a sound statistical framework when conducting market research is of paramount importance. When a business issue for investigation is decided upon, setting up a survey starts with researchers deciding which constructs are going to represent their business concepts they want to find more about (for instance, customer satisfaction).

#### 2.4.1.1 What Are Constructs?

The key ingredient in creating a survey is a *construct*. In short, constructs are quantifiable measurements that reflect a theoretical concept. The reason we use these is the often lack of quantifiability of business concepts that marketing researchers and professionals want to know more about. In our domain of business and marketing, a common construct we can take as an example is “customer satisfaction”. It is a theoretical concept that we cannot easily quantify into measurable variables. It is not possible to count customer satisfaction in particular units of measure or even assign arbitrary scores to it. However, if we generate a set of variables (which could be questions, personal attributes etc.), and if we combine these in such a way that they form a *construct*, then we have something to work with for our statistical analysis approach. Constructs are sometimes referred as *latent variables*, a denomination of variables that are inferred, via a mathematical model, from the data but they are not directly observed or measured. For the case of those constructs that are hypothesized to exist, perhaps the denomination of “*hypothetical variables*” better matches the description of a theoretical construct. This would distinguish from “*hidden variables*”, which may contain other that could be measured but, for some practical, feasibility or cost-benefit reason has not been measured so it is not part of the dataset of the study.

This means that in the analysis of some marketing datasets multiple individual variables (or measurements) are required as combined they can form a construct. It is well-documented in research covering these topics that multi-measurement items (i.e. constructs/latent variables) are better than individual measurement items as they average out the uniqueness of such individual items, make fine distinctions between people and have higher reliability [19, 75]. However, it is important when developing these measures that the domain of the construct is well specified and the items making up the multi-item construct are generated based on that domain.

Churchill [19] has created a flow diagram (Fig. 1, p. 66) with advice for creating better measurements in his 1979 publication which is still largely followed in survey research today (as shown by, for instance, Malhotra and Grover [75]). These steps are specify the domain of construct, generate sample of items, collect data, purify measure, collect data, assess reliability, assess validity and develop norms. Further,



several of these steps are feedback loops to prior steps where it may be necessary to advise changes or new inputs. He also states that the accuracy level of whether the construct has captured what intended to measure depends on the rigour with which the rules have been followed [19]. Much debate on the quality of research measures and constructs has taken place in the literature and these debates show how important it is to understand the statistical rigour required for accurate and reliable survey research.

#### 2.4.1.2 The “Cons” of Constructs

Using constructs in Psychology and Marketing research has received a lot of criticism. As stated, they represent a theoretical concept and in doing so, turn non-tangible ideas into quantifiable variables. Naturally some hiccups may be expected in this process. In the article we have discussed by Churchill [19], he refers to a publication by Jacoby in 1978 where Jacoby [58] blames much of the poor quality of marketing literature on the measures (i.e. constructs or latent variables) that researchers used to represent their theoretical items of interest. This opinion is still present today by members of the Marketing and Psychology research communities. For instance, Michell [79, p. 1] even states that “the “construct” concept is unworkable and laden with confused philosophical baggage accrued under the hegemony of logical empiricism, and its real function in psychology is obscured”. Michell concludes his highly critical and in-depth historical analysis of the construct concept by stating that as long as there is no evidence of “quantitative structure” in the theoretical attributes we call “constructs” then, there is no good scientific reason to retain the “construct” concept.

Even though the above being said, Jacoby [58], in his highly critical paper on consumer research findings states that it is of paramount importance to have *construct validity*<sup>12</sup> when conducting research using constructs. Investigating and ensuring validity in quantitative methods, specifically construct validity in survey research is not a step that can be skipped. Therefore, following Jacoby and Churchill’s arguments, there may yet be hope for survey research using constructs, as long as we follow a high standard of statistical rigour!

The main thing to remember when it comes to quantitative research is that any mistakes made during the set-up, design and execution of a research study will have confounding impacts on the reliability and validity of its findings in general as these errors at the start of the process may introduce significant flaws in the process. Measurement error is a significant problem in social sciences [4, 36]. Bagozzi et al. [4] explain that if construct validity is not assessed, the researcher cannot estimate and correct for the confounding influences of random error and method variance. This may result in ambiguous research results and possible wrongful rejection or acceptance of a hypothesis based on excessive error in measurement.

---

<sup>12</sup>[https://en.wikipedia.org/wiki/Construct\\_validity](https://en.wikipedia.org/wiki/Construct_validity).

This brings us to the statistics of creating and using constructs. Many things have been said about constructs on both sides of the argument and perhaps a lot of work still needs to be done; however, one thing is sure, survey research continues to be an integral part of Marketing, Psychology and other social sciences and ensuring its accuracy and reliability should never be ignored. Hence, we will briefly cover the basics of the vital statistics relating to constructs.

#### 2.4.1.3 The Basic Statistics of Creating Constructs

As Cronbach and Meehl [21] state, the best construct is the one around which we can build the greatest number of inferences, in the most direct fashion. Being able to do this relies on the reliability and validity of a construct. A measure (an individual variable) is *reliable* to the extent that independent but comparable measures of the same trait or construct of a given concept “agree” [19]. Basically, we can say that reliability measures the overall consistency of a measure. It is said that reliability is a necessary, but not sufficient condition, for validity. It is also true that if only single-measurement (one individual variable) is used for analysis, reliability is impossible to ascertain, hence showing that multi-measurement methods are more likely to be reliable [4].

A recommended measure for internal consistency (i.e. reliability) of a set of items (in a construct) is the coefficient alpha [19], also known as Cronbach’s  $\alpha$  or tau-equivalent reliability<sup>13</sup> [83]. Churchill states that this is the first measure a researcher should calculate as it is laden with information because the square root of coefficient alpha is the estimated correlation of the  $k$ -item test with errorless true scores. Therefore, a low alpha score indicates that the sample of items perform poorly in “capturing” the construct. Oppositely, a high alpha score means that the  $k$ -item test correlates well with true scores.

Cronbach and Meehl first coined the term “construct validity” in their 1955’s psychometric study [21]. It is almost appropriate to say that *validity* has become a sort of *holy grail* of the statistics revolving around survey research. Jacoby even states “The most necessary type of validity is construct validity” [58, p. 92]. Construct validity examines the question: Does the measure behave like the theory says a measure of that construct should behave? [21]. As stated, reliability is a necessary but not *sufficient* aspect of construct validity.

Construct validation can be done with the multitrait-multimethod (MTMM) matrix. The MTMM matrix is a correlation matrix for different concepts when each of the concepts is measured by different methods [14]. The MTMM matrix is one of the more traditional approaches for assessing construct validity and there are other methods that have been brought forward since then. For instance, Westen and Rosenthal provide two further measures of construct validation based on two effect size estimates (correlation coefficient) in their 2003 publication [106].

<sup>13</sup>[https://en.wikipedia.org/wiki/Cronbach's\\_alpha](https://en.wikipedia.org/wiki/Cronbach's_alpha).

Further, construct validity can also be evaluated through different forms of factor analysis, structural equation modelling (see Sect. 2.4.2 for more information on these methods) or other statistical measures. It is important to note however though that construct validity cannot be proven in one single research study. Rather, it is a continuous process of evaluation, reevaluation, refinement and development as the flow diagram in Fig. 1 of Churchill's 1979 paper shows [19, p. 66].

Construct validity is made up of convergent and discriminant validity. Convergent validity refers to the extent that two measures of constructs (individual variables) correlate highly with other measures designed to measure the same construct. Discriminant validity on the other hand tests whether concepts or measurements (individual variables) that are supposed to be unrelated are in fact, unrelated [4, 19]. It can also be said that discriminant validity is the extent to which the measure is actually novel and not simply a reflection of some other variable [19], (hence *unrelated* to other measures from different domains). For the keen reader who would like to learn more about survey research and the statistics of construct validation we recommend the paper by Churchill [19] or several of the other studies we have cited in this section [4, 68, 83, 106].

## 2.4.2 More Marketing Research

As we have now covered the basics of using constructs in survey research, we can go through a brief overview of common quantitative research methods used by marketing researchers. In many of these, surveys are heavily used and they provide a wide range of examples from statistical tools to decision support models and even clustering approaches. Up until recently (i.e. before the adoption of more advanced data analytics approaches from computer science disciplines), Marketing and Business researchers predominantly only used statistics analysis in their empirical research studies. Specifically, the field of multivariate statistics<sup>14</sup> is what most empirical marketing research approaches fall into. For data scientists to be able to understand where quantitative marketing researchers come from, and for the novel marketing/business analyst, this section will provide a brief overview of those methods most commonly used in marketing and consumer behaviour applications.

Multivariate statistics methods focus on the simultaneous observation and analysis of more than one outcome variable. The practical application of multivariate statistics to a particular example may involve multiple types of univariate and multivariate analyses in order to understand the relationships between variables and their relevance to the problem being studied. It is an extremely large field and includes many different statistical methods, including but not limited to, Factor Analysis, Principal Component Analysis, clustering systems, path analysis,

---

<sup>14</sup>[https://en.wikipedia.org/wiki/Multivariate\\_statistics](https://en.wikipedia.org/wiki/Multivariate_statistics).

structural equation models, etc. The book “Multivariate Data Analysis” by Hair et al. [49] provides a good starting point to learn more in-depth about these methods for those interested in doing so. We will cover the basic ideas of some of these methodologies, plus others, with a focus on those mainly used by marketing and consumer behaviour researchers and will provide directions for further reading.

#### 2.4.2.1 Factor Analysis and Principal Component Analysis

Factor analysis and Principal Component Analysis (PCA) are both well-known multivariate data analysis methods. We would like to recommend any novice reader wishing to learn more in-depth about this topic to have a look at the textbooks “Multivariate Data Analysis: A Global Perspective” by Hair et al. [48, 49] and/or “Principal Component Analysis and Factor Analysis” by Joliffe [61]. Further, for readers who are quite new to quantitative research, we recommend a “quick and dirty” introduction to factor analysis and Principal Component Analysis (PCA), via their respective Wikipedia pages.<sup>15,16</sup> As many readers may be researchers or professionals in the marketing and business analytics space and may already be familiar with these topics, we will keep it brief. We will cover the basics of Factor Analysis and PCA with a focus on their applications in the marketing domain and link to resources for further reading.

Factor analysis is a multivariate statistical technique that is concerned with the identification of structure within a set of observed variables [97]. Factor analysis is used to describe variability between observed, correlated variables in terms of a potentially lower number of unobserved variables (which are in turn called “factors”). In this description we can see the commonality with that of creating constructs from several individual measure items. In fact, some researchers use factor analysis whenever starting their analysis to determine (and reduce) the number of dimensions in their data [19]. Factor analysis essentially helps researchers minimize the number of variables in their analysis while maximizing the amount of information carried by the factors.

Factor analysis in fact can take two forms: exploratory and confirmatory factor analysis. The approach described above refers to an exploratory factor analysis where data reduction (dimensionality reduction) is the objective. Further, an exploratory factor analysis approach can be used to uncover hidden structures in the data (thanks to the dimensionality reduction). Factor analysis as an exploratory tool does not have many statistical assumptions. The only assumption is the presence of “relatedness” between the variables as represented by the correlation coefficient. If it turns out there are no correlations, then there is no underlying structure (no latent variables can be found in the individual measures). On the other hand, factor analysis can also be used as a tool to test specific hypotheses, in which case it refers

---

<sup>15</sup>[https://en.wikipedia.org/wiki/Factor\\_analysis](https://en.wikipedia.org/wiki/Factor_analysis).

<sup>16</sup>[https://en.wikipedia.org/wiki/Principal\\_component\\_analysis](https://en.wikipedia.org/wiki/Principal_component_analysis).

to a confirmatory approach [49, 97]. For instance, the measurement of customer (or user) satisfaction can be analysed in an exploratory fashion where theory perhaps can be built upon its findings. Oppositely, if the suspicion of a relationship or a set factor structures between certain measured variables and customer satisfaction already exists, then a confirmatory approach can be used to test this hypothesis [29]. For further reading about confirmatory factor analysis we recommend the book “*Confirmatory Factor Analysis for Applied Research*” by Brown [12].

Stewart [97] provides us with an interesting study on the applications, as well as the misapplications, of factor analysis in marketing. For instance, Stewart calls the use of factor analysis as a clustering technique an “extreme perversion” of the method. As factor analysis can be used to essentially “group” together independent individual variables into groups of variables (latent variables), the same technique has been used on the persons in the sample of the dataset, grouping them together instead. Hence, creating clusters of consumers, customers or whomever the population of interest may be. However, Stewart [97, p. 52] continues to explain that, “factors are not clusters” and this confusion may have arisen because a cluster is “more concrete, immediately evident and easier to understand than a factor” (a statement attributed to Raymond Cattell<sup>17</sup> in [97]).

Stewart’s 1981 paper contains an in-depth presentation of different factor analysis approaches, different methods and citations to other interesting works with marketing applications at the time. He concludes that much of the criticism that factor analysis has received has been based on misinterpretation and misunderstanding (and subsequently misapplication) of the many different methods available. The above being said, factor analysis has provided marketing and social science researchers with a useful tool to analyse many different application areas. Factor analysis has been applied in business and marketing in a wide variety of applications such as the analysis of customer-perceived service quality [98], the development and validation of a solid construct for researching business’ “innovativeness” [102], analysing destination competitiveness in a tourism application [32], analysing consumer “purchasing style” [92], customer satisfaction and relationship marketing outcomes in a banking application [59], and the analysis of social responsibility in terms of environmental marketing planning among European businesses [63]. These are a very few examples of the wide variety of applications in which factor analysis has been used. Nonetheless, they show the sheer diversity in which these long lasting statistical techniques can be applied in and help researchers analyse their respective topics of interest.

We can see that more recently, factor analysis has become a more standardized step in larger applied research studies where it is used as a construct and data validation tool in a more advanced methodological process. An example of this can be seen in Kim and Ko’s study [65] who analyse whether social media use by luxury brands enhances customer equity. A confirmatory factor analysis is used to prove the validity of each of the measurement items, before proceeding

---

<sup>17</sup>[https://en.wikipedia.org/wiki/Raymond\\_Cattell](https://en.wikipedia.org/wiki/Raymond_Cattell).

to a Structural Equation Modelling (SEM) analysis. As we learned earlier in this chapter, it is important to ensure the statistical soundness of a models' measurement constructs before deriving conclusions about a population from the study using said measurement constructs. Hence, factor analysis continues to provide a useful tool for business and marketing researchers who have extended their analytical methods well beyond the first developed factor analysis approaches. Specifically as in the example of Kim and Ko [65], SEM techniques (which include confirmatory factor analysis methods) have been used at increasingly higher rates over the last several decades with many software packages developed to help researchers in their analytic quests. More details on SEM are in the following section (Sect. 2.4.2.2).

We cover factor analysis together with Principal Component Analysis (PCA), on purpose. Factor analysis and PCA are similar and have typically been confused and used interchangeably. However, they are in fact distinctly different methods [61]. Often, PCA and factor analysis produce similar results and as Jolliffe explains, the confusion is exacerbated by software packages using the two interchangeably [61], in some cases with PCA as the default extraction method in its factor analysis routines. Many online blogs and software help pages may also help clear up the confusion between the two and provide easy-to-understand explanations for new readers in this field. One easy way of putting it is that "PCA is a linear combination of variables; Factor Analysis is a measurement model of a latent variable".<sup>18</sup>

PCA helps you to identify a new coordinate system to explore the data and the principal components bring information about the underlying structure in the data. They are the directions where there is the most variance; the directions where the data is most spread out. It provides an alternative for looking at data. Like Factor Analysis, PCA can also be described and used as a method of data reduction (or dimensionality reduction). We have to remember when statistically analysing our data, each added variable, adds a dimension. If you have, for instance, 16 variables that may be correlated you may use PCA to reduce your 16 measures to several principal components. If then, three components are extracted in this example, you may want to know the component scores (which are in fact variables added to your dataset) or you may want to look at the dimensionality of the data. If these three components accounted for 78% of the total variance, then you could say that those three dimensions in the component space account for 78% of the variance. Again, for those readers new to the field who would like to learn more in-depth about PCA, we refer them to the book by Jolliffe [61].

PCA is most often used as an exploratory analysis tool for making predictive models. PCA is not usually used to identify underlying latent variables (or constructs). However, as with its comparative method factor analysis, applications of PCA run far and wide, even within the discipline of marketing and business analytics. Some examples include an application of PCA in the analysis of organic

---

<sup>18</sup><https://www.theanalysisfactor.com/the-fundamental-difference-between-principal-component-analysis-and-factor-analysis/>.

food consumption and customer preferences towards these food products [18], using PCA as a decision model for vendor selection in a logistics application [84], an analysis of luxury brands' impact of social media marketing on customer purchase intentions and relationship marketing [66], and an analysis of online banking performance using PCA combined with data envelopment analysis [54]. Similar to factor analysis, PCA is now often used in combination with, or as part of other methodological analysis processes. Specifically some SEM Partial Least Squares (PLS) methods include PCA.

#### 2.4.2.2 Structural Equation Modelling

As stated in the previously, Structural Equation Modelling (SEM) includes methods such as confirmatory factor analysis, path analysis, partial least squares path modelling, and latent growth modelling. SEM is often used to assess latent variables and its relationships (a.k.a. “constructs” we have covered in Sect. 2.4.1.1). The SEM analysis invokes a measurement model that defines these latent variables using one or more observed variables (individual measurement items) and a structural model which imputates relationships between latent variables. Then, the links between the constructs of a structural equation model are estimated using, for instance, independent regression equations. For in-depth reading on SEM, we refer the reader to the book “Structural Equation Modelling: From Paths to Networks” by Westland [13]. Here we provide a brief history and a basic introduction of the method.

The basis for structural equation modelling was developed by the American geneticist Sewall Wright in the 1920s with his development of the statistical path analysis method [111]. Wright showed that linear relationships among observed variables can be represented in the form of so-called path diagrams and their associated path coefficients. Through tracing causal and associated paths on the diagram, the linear structural relationship between the variables was easily observed. Some 50 years later, in the 1970s, Joreskog et al. [62] developed LISREL, one of the first advanced computer programs that implement structural equation analysis which kickstarted the widespread adoption of SEM methods in the social sciences. LISREL (as its abbreviation of ‘linear structural relations’ indicates) focuses on *linear* relationships in structural equations as it is heavily based on the covariance based statistical approach initially developed by Wright. Its method is however, much more flexible and generally applicable to a wide range of models [62].

Since the development on LISREL, a plethora of software packages have been developed using a SEM method approach. For instance inside SPSS, SEM analysis is possible using AMOS,<sup>19</sup> another software that includes SEM modelling is

---

<sup>19</sup><http://www.spss.com.hk/amos/index.htm>.

Mplus<sup>20</sup> and in more recent years, Hair et al. [47] have championed the use of Partial Least Squares SEM (PLS-SEM) and the development of SMART-PLS<sup>21</sup> (more on Partial Least Squares below).

SEM first appeared in Marketing literature in the early 1980s [37] and has since become a “quasi standard in Marketing and Management research” [47]. SEM has been so widely adopted, and has been able to make a difference to such a large part of the business and marketing fields due to its ability of imputing relationships between unobserved constructs from observable variables. SEM analysis methods have been used and published in most marketing and consumer-related journals [6].

There are many applications of SEM in marketing and business literature. In first applications, the method started out as being used more as a construct validation tool [95] (like CFA). However, nowadays the method is being built upon and extended in such a way that allows researchers to build ever-more complex research models and gain greater customer insights. Besides the methodological improvements being made to SEM analysis methods and software, the range of applications in which the method is used continuously spread as well. A very quick search to some applications of SEM in the past two to three decades shows us the wide range of applications SEM is used for, even in the marketing and consumer research domains alone. Some examples of applications include; the role of commitment and trust in relationship marketing with consumers [80], analysing commitment in business relationships and partnerships [107], the analysis of customer acceptance of products in the electronics market [71] analysing the extent of customers’ willingness to repurchase based on multiple variables such as equity, value, satisfaction etc. [52], customer loyalty in retail banking [8], destination loyalty in tourism based on satisfaction and motivation [112], analysis of trust in a corporate brand [90], consumers’ perceptions of corporate social responsibility [94], the effects of social media marketing activities on customer equity in the luxury brand sector [65], and there are many applications looking at online consumer behaviour on social media, for instance, [24]. In short, the thousands (maybe hundreds of thousands) of studies conducted and published in the business, marketing and management fields that include some form of SEM analysis range widely in their applications and domains and continue to bring innovative insights to expand business knowledge.

One main new contribution to SEM analysis that we need to highlight has been the development of Partial Least Squares SEM. PLS-SEM was developed by Wold [109, 110], but has been particularly championed in recent years by Hair et al. [47] and has been widely adopted by marketing, business and psychology researchers. The reason why PLS-SEM has been so popularly received by the marketing research field is due to the fact that covariance-based SEM (CB-SEM) methods rely on a set of assumptions to be fulfilled such as, the multivariate normality of data, minimum sample size, a linear relationship between latent

---

<sup>20</sup><http://www.statmodel.com/index.shtml>.

<sup>21</sup><https://www.smartpls.com/>.



variables etc. [27, 47]. Thus, we cannot analyse non-parametric data with CB-SEM. In the cases that the assumptions necessary for CB-SEM cannot be satisfied, Partial Least Squares (PLS) analysis provides the solution as it allows non-parametric data to be analysed. Further, it is important to note when comparing CB-SEM and PLS-SEM, it is not the case that one is better than the other, but rather, different instances call for different approaches.

Hair et al. [47] further explain that the philosophical distinction between CB-SEM and PLS-SEM is as straightforward as when the research objective is theory testing and confirmation, the appropriate method would be CB-SEM. Contrary, when the research objective is prediction and theory development, then the appropriate method is PLS-SEM. Overall, when the measurement or model properties do not allow for the use of CB-SEM, or when the emphasis is more on exploration than confirmation, PLS-SEM is an attractive alternative to CB-SEM and often more appropriate [47]. We would recommend the book also written by Hair et al. titled “A primer to Partial Least Squares Structural Equation Modelling (PLS-SEM)” to any reader who would like to start their own analysis using SMART-PLS [46] or their online resources available on the SMART-PLS website.<sup>22</sup>

#### 2.4.2.3 Multiple Regression and Symbolic Regression Analysis

Conceptually and practically, PLS-SEM is similar to using multiple regression analysis. The primary objective is to maximize explained variance in the dependent constructs but additionally to evaluate the data quality on the basis of measurement model characteristics [46].

Probably all the readers are somehow familiar with simple linear regression analysis<sup>23</sup>; some of the first approaches based on least squares analysis developed in the early 1800s. Regression analysis benefits from already established statistical procedures which help to estimate the relationships among sets of variables. The focus is generally to determine the relationship between a variable (said to be “dependent” or “observed”) as a function of predictor variables (also known as independent variables). In nonlinear regression, the observed variable is modelled by a mathematical function which is a nonlinear combination of one or more predictor variables and of a set of parameters which depend on the model selected. A simple market research example is the estimation of the best fit for advertising by looking at how sales revenue (the dependent variable) changes in relation to expenditures on advertising, placement of ads, and timing of ads.

However, there are cases in which we would aim that the analysis suggest a particular model. *Symbolic regression* is a methodology in which we use computers to search a space of mathematical expressions and the aim to find the model that best fits a given set of observed data. Ideally, a mathematical model which includes many

---

<sup>22</sup><http://info.smartpls.com/index.php>.

<sup>23</sup>[https://en.wikipedia.org/wiki/Regression\\_analysis](https://en.wikipedia.org/wiki/Regression_analysis).

independent/predictor variables (and associated set of parameters) is, a priori, more likely to better approximate observed data. However, when two models are able to approximate the data with the same level of accuracy, we generally prefer the one that uses a reduced number of predictor variables and parameters. In Symbolic Regression, heuristic and metaheuristic methods of non-linear optimization have been regularly used as an approach to search for better models. In general, no particular model is provided as a starting point to the algorithm. Popular techniques are based in Genetic Programming (GP) were already discussed in the first chapter of this book and they are also applied in some other chapters as well.

These methods help to understand how the typical value of the dependent variable being investigated is affected by changes in any of the independent variables, when all other independent variables are held fixed. Far from being a “closed field”, there is a strong research focus in the area nowadays, particularly with the availability of data. Consequently, hybridization of techniques expand “simple” “old” methods hybridizing methods that have been around for a while, (e.g. PCA and multiple regression analysis [66]), or with others adaptive search techniques that are coming from the fields of computational intelligence (like Genetic Programming), or even metaheuristics, like population-based metaheuristics.

#### **2.4.2.4 Grouping and Clustering in Marketing**

As Sect. 2.2.2 covered, it is important in Customer Analytics to segment the existing market into groups of consumers that are similar, in order to be able to address their needs in a better way. However, thanks to advancements in technology, we do not have to rely only on demographics information or theory-based information alone. Instead, we can use data-driven and statistically-valid methods to segment consumers into groups in which their members are actually very similar to each other. Due to the popularity of customer segmentation, a number of methods for clustering and grouping have been applied in marketing and business analysis studies for quite a few decades.

Section II of this book covers clustering more in-depth and focuses on novel data science methods of clustering. Here, we will very briefly cover some traditional segmentation approaches more frequently used by marketing researchers prior to the adoption of more advanced data-driven analytics methods of the past two decades. First we would like to recommend the review by Punj and Stewart from 1983 [86] which covers clustering methodologies in Marketing up until that year. They focus on applications of clustering methods and the specific procedures used in clustering algorithms. Another review from a few years later by Hruschka [55] extends a segmentation review by also looking at fuzzy clustering methodologies brought forward by the literature at the time. This book covers fuzzy clustering in more detail in Chap. 3.

Two of the most common clustering techniques used traditionally in marketing literature are  $k$ -means (which is also covered in Chap. 3) and Ward’s Minimum Variance method [99]. Ward’s Minimum Variance Method is a hierarchical cluster

analysis based on Ward's 1963 proposal of an agglomerative hierarchical clustering procedure for grouping activities [103]. Since its initial publication in 1963, it has been widely used across the marketing clustering literature, built upon, further developed and implemented in much more complex methods [81]. The aim of Ward's criterion is to minimize the in-cluster variance. This means that it aims to maximize the homogeneity within a cluster. The way this works is that at each step of the method, it aims to find the pair of clusters that lead to a minimum increase in the total within-cluster variance after merging. The criterion used for the decision of whether or not to merge is based on the optimal value of an objective function which could be any objective function that accurately reflects the researchers' purpose. Ward used the error sum of squares as an example of the objective function and specifically this method is known as Ward's minimum variance method.

Another segmentation method that Punj and Stewart [86] refer to in their clustering in marketing review is iterative partitioning. This is actually a broad area of methods that differ from hierarchical clustering approaches. Iterative partitioning methods start by dividing observations into a predetermined number of clusters. Observations (objects) are then reassigned to clusters until some decision rule terminates this process. For instance,  $k$ -means is an iterative partitioning approach. With the  $k$ -means method, the pre-determined value of  $k$  indicates the number of resulting clusters. This is because the number of  $k$  sets the amount of centroids the method starts with and at the first step, each data point is assigned to its nearest centroid, based on the squared Euclidean distance. More information on the  $k$ -means method can be found in Chap. 3. Other iterative partitioning method that Punj and Stewart refer to are hill-climbing methods. With hill-climbing methods, objects are not reassigned to the cluster with the nearest centroid (as with  $k$ -means) but rather, they are moved from one cluster to another if a particular statistical criterion is obtained. Reassignment of all the objects continues until optimization occurs.

As with clustering in general, clustering methodologies used in marketing and business research take on many different forms and names, sometimes even having many different names for the same method. This makes understanding clustering or segmentation techniques used specifically in marketing research a difficult task. In order to gain an overview of the many different methods that can, and have been used for market segmentation, the review by Beane and Ennis [7] provides a good start. They include a discussion on the different clustering approaches we have covered in Sect. 2.2.2 followed by a detailed description of the many statistical and analytical methods used by marketing researchers for segmentation. They include Automatic Detection Interactor (AID), canonical analysis, factor analysis, cluster analysis, regression analysis, discriminant analysis, multidimensional scaling and conjoint analysis. These many different approaches show us that market segmentation is a hugely diverse and wide field of research in constant expansion.

Looking at more recent contributions to the literature on clustering in marketing we recommended the book by Wedel and Kamakura [105], who provide a more recent and in-depth review on segmentation approaches in marketing and cover the whole topic of segmentation from a marketing perspective, including empirical

methodologies. Further, for a look at a more topic-specific review, Dolnicar [30], focuses on data-driven segmentation methodologies used in tourism applications. Hiziroglu [53] provides us with a somewhat more technical review of clustering methodologies, focusing on soft computing approaches to customer segmentation.

Finally, in the last two decades, Finite Mixture Models (FMM) have grown in use as a clustering technique in marketing literature. Mixture models are a statistical technique that uses a probabilistic model for representing the presence of subpopulations within an overall population. Before becoming a popular method in the marketing literature, they have already been successfully applied in astronomy, biology, genetics, physics, medicine, economics and other fields [78]. FMM's are said to be "elegant procedures that incorporate mixtures of parametric distributions to define the true cluster structure" [96, 99, p. 63]. They can be used to classify observations, to adjust for clustering, and to model unobserved heterogeneity. Observed data are assumed to belong to unobserved subpopulations (clusters or classes), mixtures of probability densities or regression models are used to model the outcome of interest. When the model has been fitted, class membership probabilities can also be predicted for each observation (object). The [stata.com](http://stata.com) website provides some good introductions to finite mixture modelling and practical guides to those interested readers.<sup>24</sup> Recently, the developers of the SMART-PLS software (mentioned in Sect. 2.4.2.2), also extended their PLS algorithm to include finite mixture models to help identify and treat unobserved heterogeneity in the PLS models [46, 77]. Finally, for further reading, and a more comprehensive introduction to FFMs, we again refer the reader to the seminal work of Wedel and Kamakura [105] or the more recent updated review by Tuma et al. [99].

Next, we take a brief look at conjoint analysis in general and for further reading on clustering methodologies and new clustering ideas and algorithms, data-driven grouping methods we refer to Part II of this book.

#### 2.4.2.5 Conjoint Analysis

Conjoint analysis is a survey based statistical technique used in market research that helps determine how people value different attributes (feature, function, benefits) that make up an individual product or service. Conjoint analysis involves presenting people with choices in a survey and then analysing what the drivers for those choices are. Conjoint research approaches are very commonly used in business market research and many commercial software packages and market research services using conjoint analysis exist [43, 108]. For instance, commercial services such as Sawtooth software<sup>25</sup> provide businesses with market research services using conjoint analysis methods.

---

<sup>24</sup><https://www.stata.com/manuals/fmm.pdf>.

<sup>25</sup><https://www.sawtoothsoftware.com/>.

Conjoint analysis in social science applications was developed and championed by Green and Srinivasan and has received a lot of attention by both academic researchers and business practitioners analysing consumer behaviour since the 1970s [43, 44]. It provides a good tool for measuring a customers' trade-offs among multi-attributed products and services. In Marketing applications, conjoint analysis can be used, for example, to test customer acceptance of new product designs, to assess the appeal of advertisements and in service design.

An example of a conjoint analysis experiment would be a case where the consumer is presented with four different mutually exclusive alternatives of a product. In this case, let's take a mobile phone service plan as an example. The four options could be as follows;

- 100 min talking time, 100 texts, 5 GB data per month for \$54,
- 50 min talking time, 100 texts, 10 GB per month for \$64,
- 0 min talking time, 100 texts, unlimited GB per month for \$59, or;
- 20 min talking time, 30 texts, 20 GB per month for \$59.

The consumer would then have to (hypothetically) choose between these options which one they would take. They could also be asked to rank or rate each of the options. You may have yourself participated in such a study or taken a personality test that includes this type of design.

The aim of conjoint analysis is to be able to predict the preferences of consumers and subsequently, serve them with your offering of goods and services more effectively. From responses to the questions above, conjoint analysis uncovers the underlying value for each level, depending on how often a level was included in the product selected (each of our attributes, for instance, 100 min talking time per month). The relative value of the levels is what is relevant, in other words, how the value of one level compares to the value of another. The complete set of values that represent a consumer's trade-offs are referred to as "utilities" or "part-worths". When marketers see the part-worths, they can understand which trade-offs to make in a product or service, such that it will be most desirable to the market. This is where the true benefit of conjoint analysis and its predictive power lies.

A lot of online resources provide detailed information and practical examples for further understanding and learning of conjoint analysis' processes, advantages (and disadvantages).<sup>26,27</sup> Many different analytics and software packages easily available to academic researchers include applications for conducting statistical conjoint analysis, such as in Excel,<sup>28</sup> in SPSS<sup>29</sup> or in R programming.<sup>30</sup> For a slightly more recent academic review of conjoint analysis, we refer to Green et al.'s "30 years

---

<sup>26</sup><https://www.pragmaticmarketing.com/resources/articles/conjoint-analysis-101>.

<sup>27</sup><https://www.sawtoothsoftware.com/download/techpap/undca15.pdf>.

<sup>28</sup>[https://help.xlstat.com/customer/en/portal/articles/2062346-conjoint-analysis-in-excel-tutorial?b\\_id=9283](https://help.xlstat.com/customer/en/portal/articles/2062346-conjoint-analysis-in-excel-tutorial?b_id=9283).

<sup>29</sup><https://www.ibm.com/us-en/marketplace/spss-conjoint>.

<sup>30</sup><https://cran.r-project.org/web/packages/conjoint/conjoint.pdf>.

of conjoint analysis” paper [42], or the more recent book “Conjoint measurement: Methods and applications” by Gustafsson et al. [45]. They provide a collection of essays on conjoint analysis creating a wide and interesting view of the method as a whole, its different applications and uses in the literature.

This section is far not a conclusive list of all methods used by marketing and business researchers. It is simply a snapshot view of common methods used that heavily rely on survey research and statistical foundations and provides the reader with recommendations for further reading in those areas they may be particularly interested in.

## 2.5 Conclusion

As we have made clear, the business and marketing landscape is ever-changing and just as we may understand one aspect of consumer behaviours in one social networking platform, a new trend comes up or a new social media platform emerges. Every year we see new technologies supposedly “disrupting” the way we do business, the way we shop, communicate, search the web and even the way we live. Fortunately, however, the mathematical basis required for data analysis has been relatively well established by Applied Mathematics and Statistics since the early 1800s. It is also unequivocally stated that there is a clear hybridization of techniques in the horizon. One of the reasons is the existence of new types of hardware that can do unprecedented computations for data analysis. But another one is that problems which were previously “intractable” (i.e. those NP-complete problems discussed in the previous chapter, like the  $k$ -FEATURE SET), which are central for model building, can now be efficiently addressed with sophisticated methods based on heuristics, metaheuristics or even exact methods for small instances. This means that the new optimization techniques developed from Computer Science and Discrete Applied Mathematics will likely to have an impact in the development of Business, Customer and Data Analytics in general. The rise of Memetic Algorithms (and the whole Memetic Computing paradigm), surveyed in one of the chapters and a section of the book, is a witness of this wave of change.

The omnipresence of the internet and social networking technologies in all communications, business offerings, consumer’s product research and exchanges between business and consumers continues to increase. In this changing world, the best we can do is to continue to learn about consumers’ ever-changing tastes and apply new technologies as they emerge for the benefit of all.

In some sense, other chapters in this book try to expand on subjects presented in the previous two. Techniques in clustering, the different methods for market segmentation, the use of fuzzy logic, the use of symbolic regression, the study of techniques for graph optimization, the use of “meta-analytic” techniques, etc., are somehow extensions of the directions projected from these two chapters. Data-driven decisions are likely to be coming from a mix of different techniques, with computer science methods providing an unprecedented scalability. However,

established methods from statistics are likely to continue influencing and supporting experimental and survey design, thus we decided to present in this book some of the main “traditional” approaches as a way to create awareness to the earlier generations of computer scientists about them. We truly believe in learning from “the classics” before aiming to innovate.

It is also important for everyone coming from another discipline to get a comprehensive image of how the new technologies and capabilities available can serve business and consumers better, and how to achieve an optimal outcome for all. In our “never-ending quest for ‘absolute marketing understanding’”, marketers and business researchers need to develop more advanced analytical skills and knowledge, and the computer scientists need to learn the intricacies of marketing and business applications. The rest of this book aims to provide the readers with a wide variety of novel examples where data science, social science, marketing, consumer behaviour analysis and more are completely intertwined. We hope to tighten the “gap” between business and computer science researchers (and practitioners) and keep the conversation open and flowing!

**Acknowledgements** The authors would like to thank Dr. Jamie Carlson for his review and useful suggestions for this chapter. Pablo Moscato acknowledges previous support from the Australian Research Council Future Fellowship FT120100060 and Australian Research Council Discovery Projects DP120102576 and DP140104183.

## References

1. Sudhanshu Aggarwal, Juan A. Garay, and Amir Herzberg. Adaptive video on demand. In James H. Anderson, David Peleg, and Elizabeth Borowsky, editors, *Proceedings of the Thirteenth Annual ACM Symposium on Principles of Distributed Computing, Los Angeles, California, USA, August 14–17, 1994*, page 402. ACM, 1994.
2. Susanne Albers and Achim Passen. New online algorithms for story scheduling in web advertising. In Fedor V. Fomin, Rusins Freivalds, Marta Z. Kwiatkowska, and David Peleg, editors, *Automata, Languages, and Programming - 40th International Colloquium, ICALP 2013, Riga, Latvia, July 8–12, 2013, Proceedings, Part II*, volume 7966 of *Lecture Notes in Computer Science*, pages 446–458. Springer, 2013.
3. P. L. Andrade, G. Recalde J. Hemerly, and P. Ryan. From big data to big social and economic opportunities: Which policies will lead to leveraging data-driven innovation’s potential? *The Global Information Technology Report 2014, Geneva: World Economic Forum and INSEAD*, 2014.
4. Richard P. Bagozzi, Youjae Yi, and Lynn W. Phillips. Assessing construct validity in organizational research. *Administrative Science Quarterly*, 36(3):421–458, 1991.
5. Frank M. Bass, Douglas J. Tigert, and Ronald T. Lonsdale. Market segmentation: Group versus individual behavior. *Journal of Marketing Research*, 5(3):pp. 264–270, 1968.
6. Hans Baumgartner and Christian Homburg. Applications of structural equation modeling in marketing and consumer research: A review. *International Journal of Research in Marketing*, 13(2):139–161, 1996.
7. T. P. Beane and D. M. Ennis. Market segmentation: A review. *European Journal of Marketing*, 21(5):20–42, 1987.

8. Asunción Beerli, Josefa D. Martín, and Augustín Quintana. A model of customer loyalty in the retail banking market. *European Journal of Marketing*, 38(1/2):253–275, 2004.
9. Indranil Bose and Xi Chen. Quantitative models for direct marketing: A review from systems perspective. *European Journal of Operational Research*, 195(1):1–16, 2009.
10. Burton F. Bowman and Frederick E. McCormick. Market segmentation and marketing mixes. *Journal of Marketing*, 25(3):pp. 25–29, 1961.
11. Stephen Brown. Art or science?: Fifty years of marketing debate. *Journal of Marketing Management*, 12(4):243–267, 1996.
12. Timothy A Brown. *Confirmatory factor analysis for applied research*. Guilford Publications, 2nd edition, 2014.
13. Westland J. C. *Structural Equation Models; Studies in Systems, Decision and Control*, volume 22. Springer, Cham, 2015.
14. Donald T Campbell and Donald W Fiske. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin*, 56(2):81, 1959.
15. J. Douglas Carroll and Paul E. Green. Guest editorial: Psychometric methods in marketing research: Part i, conjoint analysis. *Journal of Marketing Research*, 32(4):385–391, 1995.
16. Gary H. Chao, Maxwell K. Hsu, and Carol Scovotti. Predicting donations from a cohort group of donors to charities: A direct marketing case study. *IJORIS*, 2(3):20–38, 2011.
17. Shiyao Chen and Lang Tong. IEMS for large scale charging of electric vehicles: Architecture and optimal online scheduling. In *IEEE Third International Conference on Smart Grid Communications, SmartGridComm 2012, Tainan, Taiwan, November 5–8, 2012*, pages 629–634. IEEE, 2012.
18. Gaetano Chinnici, Mario D’Amico, and Biagio Pecorino. A multivariate statistical analysis on the consumers of organic products. *British Food Journal*, 104(3/4/5):187–199, 2002.
19. Gilbert A. Churchill. A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, 16(1):64–73, 1979.
20. Tim Cooper, Ryan LaSalle, and Kuangyi Wei. Guarding and growing personal data value. Technical report, Accenture, 01 2016.
21. L. J. Cronbach and P. E. Meehl. Construct validity in psychological tests. *Psychological Bulletin*, 52(4):281–302, 1955.
22. Marek Cygan, Marcin Mucha, Piotr Sankowski, and Qiang Zhang. Online pricing with impatient bidders. In Robert Krauthgamer, editor, *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10–12, 2016*, pages 190–201. SIAM, 2016.
23. Natalie J. de Vries, Ahmed Shamsul Arefin, and Pablo Moscato. Gauging heterogeneity in online consumer behaviour data: A proximity graph approach. In *IEEE Fourth International Conference on Social Computing and Big Data and Cloud Computing*. IEEE, 2014.
24. Natalie J. de Vries and Jamie Carlson. Customer engagement, brands and social media. *Journal of Brand Management*, 21(6):495–515, Aug 2014.
25. Natalie J. de Vries, Jamie Carlson, and Pablo Moscato. A data-driven approach to reverse engineering customer engagement models: Towards functional constructs. *PLOS ONE*, 9(7):e102768, 2014.
26. Joel Dean. Problems of product-line pricing. *Journal of Marketing*, 14(4):518–528, 1950.
27. Adamantios Diamantopoulos and Judy A. Siguaw. *Introducing LISREL*. Sage Publishing, 2000.
28. Peter R. Dickson and James L. Ginter. Market segmentation, product differentiation, and marketing strategy. *Journal of Marketing*, 51(2):pp. 1–10, 1987.
29. William J. Doll, Weidong Xia, and Gholamreza Torkzadeh. A confirmatory factor analysis of the end-user computing satisfaction instrument. *MIS Quarterly*, 18(4):453–461, 1994.
30. Sara Dolnicar. A review of data-driven market segmentation in tourism. *Journal of Travel & Tourism Marketing*, 12(1):1–22, 2002.
31. Peter F. Drucker. Marketing and economic development. *Journal of Marketing*, 22(3):pp. 252–259, 1958.



32. Larry Dwyer, Robert Mellor, Zelko Livaic, Deborah Edwards, and Chulwon Kim. Attributes of destination competitiveness: a factor analysis. *Tourism analysis*, 9(1-1):91–101, 2004.
33. A. S. C. Ehrenberg. The pattern of consumer purchases. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 8(1):26–41, 1959.
34. Sunil Erevelles, Nobuyuki Fukawa, and Linda Swayne. Big data consumer analytics and the transformation of marketing. *Journal of Business Research*, “Article in Press”:8 pages, 2015.
35. Jon Feldman, Monika Henzinger, Nitish Korula, Vahab S. Mirrokni, and Clifford Stein. Online stochastic packing applied to display ad allocation. In Mark de Berg and Ulrich Meyer, editors, *Algorithms - ESA 2010, 18th Annual European Symposium, Liverpool, UK, September 6–8, 2010. Proceedings, Part I*, volume 6346 of *Lecture Notes in Computer Science*, pages 182–194. Springer, 2010.
36. D. W. Fiske. Convergent-discriminant validation in measurements and research strategies. *New Directions for Methodology of Social & Behavioral Science*, 12:77–92, 1982.
37. Claes Fornell and David F. Larcker. Structural equation models with unobservable variables and measurement error: Algebra and statistics. *Journal of Marketing Research*, 18(3):382–388, 1981.
38. Ronald E. Frank and William F. Massy. Market segmentation and the effectiveness of a brand’s price and dealing policies. *The Journal of Business*, 38(2):pp. 186–200, 1965.
39. Negin Golrezaei, Hamid Nazerzadeh, and Paat Rusmevichientong. Real-time optimization of personalized assortments. *Management Science*, 60(6):1532–1551, 2014.
40. Sulekha Goyat. The basis of market segmentation: a critical review of literature. *European Journal of Business and Management*, 3(9):45–54, 2011.
41. Paul E. Green. A new approach to market segmentation. *Business Horizons*, 20(1):61–73, 1977.
42. Paul E. Green, Abba M. Krieger, and Yoram Wind. Thirty years of conjoint analysis: Reflections and prospects. *Interfaces*, 31(3\_supplement):S56–S73, 2001.
43. Paul E. Green and V. Srinivasan. Conjoint analysis in marketing: New developments with implications for research and practice. *Journal of Marketing*, 54(4):3–19, 1990.
44. Paul E Green and Venkatachary Srinivasan. Conjoint analysis in consumer research: issues and outlook. *Journal of consumer research*, 5(2):103–123, 1978.
45. Anders Gustafsson, Andreas Herrmann, and Frank Huber. *Conjoint measurement: Methods and applications*. Springer Science & Business Media, 2013.
46. Joe F. Hair, G. Tomas M. Hult, Christian M. Ringle, and Marko Sarstedt. *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. Sage Publishing, 2 edition, 2016.
47. Joe F. Hair, Christian M. Ringle, and Marko Sarstedt. PLS-sem: Indeed a silver bullet. *Journal of Marketing Theory and Practice*, 19(2):139–152, 2011.
48. Joseph F Hair, William C Black, Barry J Babin, and Anderson. *Multivariate Data Analysis*. Prentice hall Upper Saddle River, NJ, 7 edition, 2009.
49. Joseph F Hair, William C Black, Barry J Babin, Rolph E Anderson, Ronald L Tatham, et al. *Multivariate data analysis*, volume 5. Prentice hall Upper Saddle River, NJ, 1998.
50. Russell I. Haley. Benefit segmentation: A decision-oriented research tool. *Journal of Marketing*, 32(3):pp. 30–35, 1968.
51. Andrew Hasty. Treating consumer data like oil: How re-framing digital interactions might bolster the federal trade commission’s new privacy framework notes. *Federal Communications Law Journal*, 67:293, 2014–2015.
52. Phillip K. Hellier, Gus M. Geursen, Rodney A. Carr, and John A. Rickard. Customer repurchase intention: A general structural equation model. *European Journal of Marketing*, 37(11/12):1762–1800, 2003.
53. Abdulkadir Hiziroglu. Soft computing applications in customer segmentation: State-of-art review and critique. *Expert Systems with Applications*, 40(16):6491–6507, 2013.
54. Chien-Ta Bruce Ho and Desheng Dash Wu. Online banking performance evaluation using data envelopment analysis and principal component analysis. *Computers & Operations Research*, 36(6):1835–1842, 2009.

55. H. Hruschka. Market definition and segmentation using fuzzy clustering methods. *International Journal of Research in Marketing*, 3(2):117–134, 1986.
56. Xinhui Hu, Arne Ludwig, Andréa W. Richa, and Stefan Schmid. Competitive strategies for online cloud resource allocation with discounts: The 2-dimensional parking permit problem. In *35th IEEE International Conference on Distributed Computing Systems, ICDCS 2015, Columbus, OH, USA, June 29–July 2, 2015*, pages 93–102. IEEE, 2015.
57. Philip Hunter. Biology is the new Physics. *EMBO Report*, 11(5):350–352, 2010.
58. Jacob Jacoby. Consumer research: A state of the art review. *Journal of Marketing*, 42(2):87–96, 1978.
59. Vimi Jham and Kaleem Mohd Khan. Determinants of performance in retail banking: Perspectives of customer satisfaction and relationship marketing. *Singapore Management Review*, 30(2):35, 2008.
60. Richard M. Johnson. Market segmentation: A strategic management tool. *Journal of Marketing Research*, 8(1):pp. 13–18, 1971.
61. Ian T Jolliffe. Principal component analysis and factor analysis. In *Principal component analysis*, pages 115–128. Springer, 1986.
62. Karl G Jöreskog and Marielle Thiilo. Lisrel a general computer program for estimating a linear structural equation system involving multiple indicators of unmeasured variables. *ETS Research Report Series*, 1972(2), 1972.
63. Jari Kärnä, Eric Hansen, and Heikki Juslin. Social responsibility in environmental marketing planning. *European Journal of Marketing*, 37(5/6):848–871, 2003.
64. Richard M. Karp. On-line algorithms versus off-line algorithms: How much is it worth to know the future? In Jan van Leeuwen, editor, *Algorithms, Software, Architecture - Information Processing '92, Volume 1, Proceedings of the IFIP 12th World Computer Congress, Madrid, Spain, 7–11 September 1992*, volume A-12 of *IFIP Transactions*, pages 416–429. North-Holland, 1992.
65. Angella J. Kim and Eunju Ko. Do social media marketing activities enhance customer equity? An empirical study of luxury fashion brand. *Journal of Business Research*, 65(10):1480–1486, 2012. Fashion Marketing and Consumption of Luxury Brands.
66. Angella Jiyoung Kim and Eunju Ko. Impacts of luxury fashion brand's social media marketing on customer relationship and purchase intention. *Journal of Global Fashion Marketing*, 1(3):164–171, 2010.
67. Philip Kotler. Behavioral models for analyzing buyers. *Journal of Marketing*, 29(4):pp. 37–45, 1965.
68. Jon A. Krosnick. Survey research. *Annual Review of Psychology*, 50(1):537–567, 1999. PMID: 15012463.
69. V. Kumar. Evolution of marketing as a discipline: What has happened and what to look out for. *Journal of Marketing*, 79(1):1–9, 2015.
70. Jack A. Lesser and Marie Adele Hughes. The generalizability of psychographic market segments across geographic locations. *Journal of Marketing*, 50(1):pp. 18–27, 1986.
71. Ting-Peng Liang and Jin-Shiang Huang. An empirical study on consumer acceptance of products in electronic markets: a transaction cost model. *Decision Support Systems*, 24(1):29–43, 1998.
72. Barak Libai, Ruth Bolton, Marnix S. Bügel, Ko de Ruyter, Oliver Götz, Hans Risselada, and Andrew T. Stephen. Customer-to-customer interactions: Broadening the scope of word of mouth research. *Journal of Service Research*, 13(3):267–282, 2010.
73. Christopher Lovelock and Evert Gummesson. Whither services marketing?: In search of a new paradigm and fresh perspectives. *Journal of Service Research*, 7(1):20–41, 2004.
74. Christopher Lovelock, Paul Patterson, and Jochen Wirtz. *Services Marketing*. Pearson Australia Group Pty Ltd, 6 edition, 2015.
75. Manoj K Malhotra and Varun Grover. An assessment of survey research in POM: from constructs to theory. *Journal of Operations Management*, 16(4):407–425, 1998.
76. William F. Massy and Ronald E. Frank. Short term price and dealing effects in selected market segments. *Journal of Marketing Research*, 2(2):pp. 171–185, 1965.

77. Lucy M Matthews, Marko Sarstedt, Joseph F Hair, and Christian M Ringle. Identifying and treating unobserved heterogeneity with fimix-pls: Part ii - a case study. *European Business Review*, 28(2):208–224, 2016.
78. Geoffrey McLachlan and David Peel. *Finite mixture models*. John Wiley & Sons, 2004.
79. Joel Michell. Constructs, inferences, and mental measurement. *New Ideas in Psychology*, 31(1):13–21, 2013. On defining and interpreting constructs: Ontological and epistemological constraints.
80. Robert M. Morgan and Shelby D. Hunt. The commitment-trust theory of relationship marketing. *Journal of Marketing*, 58(3):20–38, 1994.
81. Fionn Murtagh and Pierre Legendre. Ward’s hierarchical agglomerative clustering method: Which algorithms implement ward’s criterion? *Journal of Classification*, 31(3):274–295, Oct 2014.
82. T. A. Oliveira, Vitor N. Coelho, Marcone J. F. Souza, D. L. T. Boava, F. Boava, Igor Machado Coelho, and Bruno N. Coelho. A hybrid variable neighborhood search algorithm for targeted offers in direct marketing. *Electronic Notes in Discrete Mathematics*, 47:205–212, 2015.
83. J Paul Peter. Reliability: A review of psychometric basics and recent marketing practices. *Journal of marketing research*, pages 6–17, 1979.
84. Alberto Petroni and Marcello Braglia. Vendor selection using principal component analysis. *Journal of Supply Chain Management*, 36(1):63–69, 2000.
85. Joseph T. Plummer. The concept and application of life style segmentation. *Journal of Marketing*, 38(1):pp. 33–37, 1974.
86. Girish Punj and David W. Stewart. Cluster analysis in marketing research: Review and suggestions for application. *Journal of Marketing Research*, 20(2):134–148, 1983.
87. Alan A Roberts. Applying the strategy of market segmentation. *Business Horizons*, 4(3):65–72, 1961.
88. Roland T. Rust and Ming-Hui Huang. The service revolution and the transformation of marketing science. *Marketing Science*, 33(2):206–221, 2014.
89. Leon Schiffman, Aron O’Cass, Angela Paladino, and Jamie Carlson. *Consumer Behaviour*. Pearson Australia, Frenchs Forest NSW, 6 edition, 2014.
90. Christina Sichtmann. An analysis of antecedents and consequences of trust in a corporate brand. *European Journal of Marketing*, 41(9/10):999–1015, 2007.
91. A. Prasad Sistla, Ouri Wolfson, Yelena Yesha, and Robert H. Sloan. Towards a theory of cost management for digital libraries and electronic commerce. *ACM Trans. Database Syst.*, 23(4):411–452, 1998.
92. Noel Y. M. Siu, Charlie C. L. Wang, Ludwig M. K. Chang, and Alice S. Y. Hui. Adapting consumer style inventory to chinese consumers. *Journal of International Consumer Marketing*, 13(2):29–47, 2001.
93. Wendell R. Smith. Product differentiation and market segmentation as alternative marketing strategies. *Journal of Marketing*, 21(1):3–8, 1956.
94. Andrea J. S. Stanaland, May O. Lwin, and Patrick E. Murphy. Consumer perceptions of the antecedents and consequences of corporate social responsibility. *Journal of Business Ethics*, 102(1):47–55, Aug 2011.
95. Jan-Benedict E.M. Steenkamp and Hans C.M. van Trijp. The use of lisrel in validating marketing constructs. *International Journal of Research in Marketing*, 8(4):283–299, 1991.
96. Douglas Steinley and Michael J Brusco. Evaluating mixture modeling for clustering: Recommendations and cautions. *Psychological Methods*, 16(1):63, 2011.
97. David W. Stewart. The application and misapplication of factor analysis in marketing research. *Journal of Marketing Research*, 18(1):51–62, 1981.
98. GS Sureshchandar, Chandrasekharan Rajendran, and RN Anantharaman. Determinants of customer-perceived service quality: a confirmatory factor analysis approach. *Journal of services Marketing*, 16(1):9–34, 2002.
99. Michael N Tuma, Reinhold Decker, and Sören W Scholz. A survey of the challenges and pitfalls of cluster analysis application in market segmentation. *International Journal of Market Research*, 53(3):391–414, 2011.

100. Jenny van Doorn, Katherine N. Lemon, Vikas Mittal, Stephan Nass, Doreén Pick, Peter Pirner, and Peter C. Verhoef. Customer engagement behavior: Theoretical foundations and research directions. *Journal of Service Research*, 13(3):253–266, 2010.
101. Donald E. Vinson, Jerome E. Scott, and Lawrence M. Lamont. The role of personal values in marketing and consumer behavior. *Journal of Marketing*, 41(2):pp. 44–50, 1977.
102. Catherine L. Wang and Pervaiz K. Ahmed. The development and validation of the organisational innovativeness construct using confirmatory factor analysis. *European Journal of Innovation Management*, 7(4):303–313, 2004.
103. Joe H Ward Jr. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301):236–244, 1963.
104. Michael Wedel and Wagner A. Kamakura. *Market Segmentation: Conceptual and Methodological Foundations*, volume 8 of *International Series in Quantitative Marketing*. Springer, New York, USA, 6 edition, 2000.
105. Michel Wedel and Wagner A Kamakura. *Market segmentation: Conceptual and methodological foundations*, volume 8. Springer Science & Business Media, 2000.
106. Drew Westen and Robert Rosenthal. Quantifying construct validity: two simple measures. *Journal of personality and social psychology*, 84(3):608, 2003.
107. Martin Wetzels, Ko de Ruyter, and Marcel van Birgelen. Marketing service relationships: the role of commitment. *Journal of Business & Industrial Marketing*, 13(4/5):406–423, 1998.
108. Dick R Wittink and Philippe Cattin. Commercial use of conjoint analysis: An update. *The Journal of Marketing*, pages 91–96, 1989.
109. H. Wold. Systems analysis by partial least squares. In P. Nijkamp, L. Leitner, and N. Wrigley, editors, *Measuring the Unmeasurable*, pages 221–225. Marinus Nijhoff, 1985.
110. Herman Wold. Causal flows with latent variables: partings of the ways in the light of nipals modelling. *European Economic Review*, 5(1):67–86, 1974.
111. Sewall Wright. Correlation and causation. *Journal of agricultural research*, 20(7):557–585, 1921.
112. Yooshik Yoon and Muzaffer Uysal. An examination of the effects of motivation and satisfaction on destination loyalty: a structural model. *Tourism Management*, 26(1):45–56, 2005.