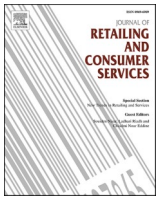
Journal of Retailing and Consumer Services 67 (2022) 102989

Contents lists available at ScienceDirect

Journal of Retailing and Consumer Services

journal homepage: www.elsevier.com/locate/jretconser



Mining the text of online consumer reviews to analyze brand image and brand positioning

Miriam Alzate *, Marta Arce-Urriza, Javier Cebollada

Public University of Navarra, Pamplona, Spain

ARTICLE INFO

Keywords:
eWOM
Online reviews
Text mining
Brand positioning

ABSTRACT

The growth of the Internet has led to massive availability of online consumer reviews. So far, papers studying online reviews have mainly analysed how non-textual features, such as ratings and volume, influence different types of consumer behavior, such as information adoption decisions or product choices. However, little attention has been paid to examining the textual aspects of online reviews in order to study brand image and brand positioning. The text analysis of online reviews inevitably raises the concept of "text mining"; that is, the process of extracting useful and meaningful information from unstructured text. This research proposes an unified, structured and easy-to-implement procedure for the text analysis of online reviews with the ultimate goal of studying brand image and brand positioning. The text mining analysis is based on a lexicon-based approach, the Linguistic Inquiry and Word Count (Pennebaker et al., 2007), which provides the researcher with insights into emotional and psychological brand associations.

1. Introduction

Brand positioning is a crucial step in marketing strategy. Markets are overcrowded with products, and, to simplify the buying process, consumers organize products into categories and position them in their minds (Kotler and Armstrong, 2020). The position of a brand is formed by a complex set of consumer perceptions, images, and emotions associated with the brand's products and how they compare with competing products. Therefore, to position a product, companies need to understand how consumers perceive products in its category. Traditionally, companies gathered information by surveying consumers. However, with the growth of the Internet and social media, there is an overwhelming amount of online information for both consumers and firms. A particular way in which consumers share information about companies, products and brands is through electronic word-of-mouth (eWOM). Whereas there is a vast literature on the ways consumers generate and use this information in their purchasing decisions (Chevalier and Mayzlin, 2006; Hofmann et al., 2017), less attention has been paid to the insights firms can draw from it (Verma and Yadav, 2021), for purposes such as brand positioning for example.

One of the main types of eWOM are online consumer reviews. Others include comments on social networks (e.g., posts on Facebook and tweets on Twitter), comments in forums, blog entries, etc. Online consumer reviews mainly comprise text (Berger et al., 2020), and can be defined as any "positive, neutral, or negative evaluation of a product, service, person, or brand presumed to be posted by former customers on websites that host consumer reviews" (Filieri et al., 2018). By exploring the text of online reviews, companies have the opportunity to expand and deepen their knowledge of aspects such as consumer preferences, brand image, brand associations, and brand positioning (Balducci and Marinova, 2018; Hartmann et al., 2019; Kübler et al., 2019). Until now, the qualitative nature of the text content of online reviews has made it difficult to analyze at an aggregate level (e.g., product, brand or company-level) and to extract meaningful insights. (Chen et al., 2015). However, the development of high-speed computers has enabled companies and researchers to advance a step further in the study of texts and language (Chung and Pennebaker, 2007; Tausczik and Pennebaker, 2010).

The various available text mining techniques for the exploration of online reviews are usually classified into machine learning and lexicon-based methods (Hartmann et al., 2019; Kübler et al., 2019). The selection of one method over another depends on two key factors: the specific research objectives of the firm and the skills required to conduct the analysis. Machine learning algorithms require a level of expertise computational skills beyond the usual capacity possibilities of small and medium firms. A report by Magoulas and Swoyer (2020) along these

* Corresponding author. Campus de Arrosadia s/n, 31006, Pamplona (Navarra), Spain.
E-mail addresses: miriam.alzate@unavarra.es (M. Alzate), marta.arce@unavarra.es (M. Arce-Urriza), cebollada@unavarra.es (J. Cebollada).

https://doi.org/10.1016/j.jretconser.2022.102989
Received 26 November 2021; Received in revised form 31 January 2022; Accepted 14 March 2022
Available online 14 April 2022

lines reveals that one of the main reasons why companies do not adopt further Artificial Intelligence (AI) algorithms, such as machine learning, is the "lack of skilled personnel/difficulty in filling the required roles". On the other hand, lexicon-based methods, which rely on established dictionaries of words, provide a simpler and more intuitive means to analyze the text of online reviews, which is better suited to the possibilities of small and medium companies. The lexicon-based tool Linguistic Inquiry and Word Count (LIWC) developed by Pennebaker et al. (2007) is used for the text mining analysis in this research.

This research is practical in nature and contributes to existing literature by suggesting a procedure for the analysis of the textual content of online reviews in the context of brand image and brand positioning. Most research papers have collected brand image data using quantitative techniques, in particular, surveys (Cho et al., 2015; Davis et al., 2009; John et al., 2006; Kim et al., 2003; Low and Lamb, 2000; Malhotra, 1981; Park and Rabolt, 2009), or qualitative techniques, in general in depth interviews, to elicit brand associations and build Brand Concept Maps (BCM). However, the wide availability of consumer opinions in the digital environment, together with the increasing number of tools for the analysis of unstructured data, as contained in online texts, has shifted attention to the study of brand image using the language and narrative used in eWOM texts (Culotta and Cutler, 2016; Heng et al., 2018; Hu and Trivedi, 2020; Netzer et al., 2012).

There are advantages in the use of eWOM data over survey data or data obtained from in-depth interviews. One of the most relevant characteristics of eWOM is its spontaneity (Marchand et al., 2017; Yang and Cho, 2015), in that consumers are more likely to be truthful in expressing their brand perceptions and reporting their own behavior. It is generated without direct prompting or influence from marketers, and is usually motivated by a desire to help, warn, or communicate status to others (Kozinets et al., 2010). Moreover, in contrast to surveys and in-depth interviews, which provide primary information, eWOM is a secondary source of information that is widely available online, so it is quite easy to collect huge amounts of online opinions automatically, using techniques such as web scraping and social networks APIs. Besides, as claimed by Bijmolt et al. (2021), questionnaires used to evaluate brand perceptions are generally extensive and involve respondents in time-consuming judgment tasks. Nevertheless, eWOM also has its disadvantages. First, the huge quantity of text could expose firms and scholars to information overload; and second, the analysis of eWOM texts is a relatively new domain, requiring knowledge and expertise that companies and scholars do not usually possess. Overall, one of the main advantages of using eWOM to study brand image and positioning is that, because eWOM occurs without prompting, consumers are more likely to convey their true product and brand perceptions, while being less likely to lie or succumb to potential social desirability in the presence of an interviewer. Moreover, consumers writing online reviews can say anything they like, so brand image researchers have no need to keep to a given scale or interview structure. This might allow them to discover brand image associations that might not emerge from a structured questionnaire or interview.

Our proposal for the analysis of online reviews is the Linguistic Inquiry and Word Count (LIWC) because it provides a wide range of psychological process variables. Most of the existing literature using online reviews to explore brand image rely on the analysis of product or brand features, which are more closely related to physical attributes. Nevertheless, the brand image literature has pointed out the relevancy of also considering the psychological benefits of brands (Delgado-Ballester and Fernández-Sabiote, 2016; Ruane and Wallace, 2015; Simon et al., 2016). Thus, the use of the LIWC allows us to uncover and analyze the psychological rewards that consumers associate with brands.

In terms of managerial implications, the main contribution of this research is to provide a procedure that is structured, unified and easy-to-follow, in order to enable small and medium companies to understand brand associations and build positioning maps using data from online consumer reviews. As claimed by Veloutsou and Delgado-Ballester (2018), thanks to increasing use of the Internet, individuals can easily voice their assessments and feelings about brands and reach a wide audience through online platforms and social media. Therefore, for their brand to prosper, it is vital that companies improve their skill at handling such messages.

By "structured procedure" we mean one with clearly marked out stages for brand positioning analysis based entirely on the text mining of online reviews to extract brand associations. We describe the procedure as "unified" because it combines concepts derived from a review of the text mining and brand positioning literatures in a single study, to produce an easy-to-follow guide or procedure involving techniques from both areas. We claim that it is an "easy-to-follow procedure" because the text mining method, in this case, the LIWC, is lexicon-based, inexpensive and intuitive, since it requires no knowledge of machine learning. The user simply needs to input the text files for analysis to automatically obtain data on more than 90 text variables, including psychological associations. After the text mining analysis, two commonly-used techniques are applied in the brand positioning analysis: Principal Component Analysis to build a perceptual map and Hierarchical Clustering to identify brand subgroups. In this case, these tasks are conducted in R software, which is available free of charge both to private individuals and to companies. For company use, there exist R software packages specifically-created for brand positioning analysis.

To illustrate our proposal, we provide an empirical application, in which we use the entire set of online consumer reviews for a category of cosmetic products (blushers) available from a popular US online cosmetics retailer on February the 17th 2017. A total of 62,496 online reviews on 44 different cosmetics brands was analysed. The blusher category was chosen for one main reason. Cosmetics are classified by Girard and Dion (2010) as experience products, which Nelson (1970) defines as those whose attribute information cannot be known before use or consumption, unlike search products, whose attribute information (e.g., price, quality, size, and dimension) can be easily evaluated prior to consumption. Since experience products are so difficult to evaluate, consumers usually rely more heavily on recommendations than in the case of search products, for which they are more likely to use their own decision-making processes (King and Balasubramanian 1994; Senecal and Nantel, 2004). When it comes to exploring brand image, therefore, retail managers might find an experiential category even more relevant than a search product category.

The remainder of this paper is structured as follows. In section 2, we provide an overview of the literature on brand associations, brand image and brand positioning, and existing approaches to the measurement of these concepts. We also review the literature on text-mining approaches and tools for discovering emotions and topics from text. Section 3 presents the proposed research procedure for exploring brand positioning using online consumer reviews. The results of the analyses are presented in Section 4. Section 5 contains a discussion of the findings, including an analysis of the main managerial implications and remarks highlighting the main limitations and areas for future research.

2. Literature review

The literature review is split into three main areas: the concept and relevance of brand image and brand positioning for marketing strategy; previous brand image and brand positioning measurement methods; and the literature on the use of text mining techniques to identify brand image from consumer-generated text content.

2.1. The concepts of brand image and brand positioning

To understand and contextualize the relevance of brand image and brand positioning, we have to go back to the brand equity literature, pioneered by Aaker (1991), who defined brand associations as "brands assets and liabilities that include anything linked in memory to a brand". John et al. (2006) posit that consumers might associate a brand with

attributes, features and usage situations. The set of brand associations together form a brand image. Both brand image and brand associations represent what the brand means to consumers. This meaning usually develops from the consumer's own experience with the brand, the firm's marketing mix activities (Aaker, 1991) and the opinions of other consumers. Brand associations are important; both to companies and to consumers (Low and Lamb, 2000). Companies use them to foster positive attitudes towards the brand, suggest associated its benefits and position it in the marketplace, while consumers use them to process, organize and retrieve information from their memory store to aid purchase decision making (Aaker, 1991). Overall, scholars have found that favorable brand image and brand attitudes have a positive impact on purchase intentions (Baksi and Panda, 2018; Jalilvand and Samiei, 2012; Kudeshia and Kumar, 2017; Spears and Singh, 2004). According to Henderson et al. (1998), brand associations that evoke positive affect, as well as cognitive considerations of benefits, provide consumers with reasons for buying a brand or product.

Another important way in which brand associations create value to the firm is by providing a basis for brand and product positioning (Aaker, 1991). The positioning of the product or brand can be defined as the place it occupies relative to competing products/brands in consumers' minds (Kotler and Armstrong, 2020). This position is based on the key associations (e.g., attributes and benefits) used by consumers to evaluate the various alternatives available in the market. To explore brand positioning, companies can build perceptual positioning maps, which show consumer perceptions of their own brands versus those of their competitors with respect to key purchasing dimensions (Kotler and Armstrong, 2020). When evaluating the position of brands in the marketplace, companies are also dealing with a brand competition analysis. As claimed by France and Ghose (2016), analyzing brand competition and identifying sets of competing brands is a core marketing activity. Market competition for a product category can be described as a set of product submarkets, where each brand belongs to a submarket and brands within a submarket compete with one another (France and Ghose, 2016). Submarkets can be defined not only on the basis of product features, which might be easily quantifiable attributes such as price, but also of less quantifiable attributes, such as consumer perceptions (France and Ghose, 2016; Urban et al., 1984).

Veloutsou and Delgado-Ballester (2018) point out that one of the challenging areas of branding research is brand co-creation between the company and external agents, such as consumers. Consumers, among other agents in today's online marketing context, are gaining a more powerful voice in their role as eWOM generators and active participants in the communication of brand emotions and associations to other consumers. Therefore, the analysis of online reviews, a specific type of eWOM, is seen as a key value-adding factor in the branding literature.

2.2. The measurement of brand image and brand positioning

2.2.1. The use of quantitative techniques

Most research studies in the literature have used surveys to approach brand image as a multidimensional concept, and measured it with pre-existing scales in the literature (Anselmsson et al., 2017; Swoboda et al., 2016; Baksi and Panda, 2018; Bhat and Chakraborty1, 2018; Cho et al., 2015; Davis et al., 2009; DeSarbo et al., 2011; John et al., 2006; Kim et al., 2003; Konuk, 2018; Londoño et al., 2016; Low and Lamb, 2000; Malhotra, 1981; Panda et al., 2019; Park and Rabolt, 2009).

Some papers use general brand image scales, not specifically adapted to the product category. For example, Martínez Salinas and Pina Pérez (2009) analyze the effect of brand extensions on brand image, using a brand image scale adapted from previous studies. Using this scale, the authors evaluate functional image (e.g., "the products have a high quality"), affective image (e.g., "the brand is nice") and reputation (e.g., "It is one of the best brands in the sector"). There is also a stream of literature that considers brand image as being related to the product category within which the brand is marketed (Baksi and Panda, 2018; Low and Lamb, 2000). For example, Baksi and Panda (2018) analyze the destination image of several cities in India using a survey including 43 dimensions adapted from scales in previous studies (e.g., "safe and secure environment", "entertainment in festivals" and "physical landscape of the destination").

Overall, the brand image literature advocates that, when measuring brand associations, it is necessary to consider not only physical attributes, but also functional, emotional and self-expressive benefits (Büyükdağ and Kitapci, 2021; Delgado-Ballester and Fernández-Sabiote, 2016; Ruane and Wallace, 2015; Simon et al., 2016). As claimed by Veloutsou and Delgado-Ballester (2018), consumers use and buy brands to gain certain psychological rewards. In this sense, they try to satisfy their functional, emotional, personal and social needs through the value offered by the brand.

2.2.2. The use of qualitative techniques

An important stream of research has used qualitative data collection techniques, particularly in-depth interviews. In general, these studies rely on the construction of Brand Concept Maps (BCM) to analyze brand image and brand positioning. BCM are based on the assumption that embedded in their structure one can find the inherent content (concepts and their associations) and relationships (links between concepts and associations) represented in consumers' minds (Brandt et al., 2011).

There are three stages in the BCM construction process: data elicitation, the representation of data as graph-theoretical or spatial structures, and the application of network analytic techniques (Henderson et al., 1998). Several techniques are used during in-depth interviews for the elicitation of brand associations, ranging from the highly qualitative free association and free response techniques (Cheng-Hsui Chen, 2001; Olson and Muderrisoglu, 1979) to the more structured repertory grid technique (Chang and Mak, 2018; Hu and Trivedi, 2020; Kawaf and Istanbulluoglu, 2019; Kelly, 1991; Whyte, 2018), laddering (Park et al., 2019; Reynolds and Gutman, 1979; Rossolatos, 2019) and pairwise similarities (Hauser and Koppelman, 1979; Teichert et al., 2017). All these techniques have been used not only to explore brand image and brand positioning but also to analyze consumer motivations and other facets of consumer behavior.

As Stated by Brandt et al. (2011), one of the disadvantages of quantitative methods (compared with qualitative techniques) is that they place the emphasis on the conscious processes in brand evaluation, while qualitative techniques enable the researcher, in addition, to elicit "hidden" or unconscious information. For a better analysis of brand image and brand positioning, the authors recommend using the combined strengths of both methods.

In this research, the analysis of online review texts allows us to combine the strengths of qualitative and quantitative techniques. On the one hand, the text analysis is qualitative, since consumers express their feelings, motivations and experiences through their written opinions. On the other hand, the availability of huge quantities of online reviews, together with the proliferation of text mining techniques and statistical methods, allows researchers to benefit from the advantages of quantitative techniques for analyzing brand image and positioning from the text of online reviews.

2.3. Mining the text of online reviews to explore brand image and brand positioning

The study of eWOM text content offers scholars and companies a great opportunity to deepen their knowledge of brand image. Narrative and persuasive language have been widely examined in several research domains, such as communication, psychology and marketing (Areni, 2003; Hamby et al., 2015; Holtgraves and Lasky, 1999; Li et al., 2019). Tausczik and Pennebaker (2010) claim that language is the way in which people express their internal thoughts and emotions. Consistent with this idea, psychologists have found that personality traits can be gleaned from linguistic cues, including aspects such as topics discussed,

style, syntax, lexicon and type of speech (Walker et al., 2007).

Although the study of narrative and persuasion has its roots outside the digital environment, the concept of "text mining" has emerged to refer to the process of extracting useful and meaningful information from large amounts of text that can be found online (Netzer et al., 2012). Text mining comprises a set of techniques and technologies that are used to explore large amounts of text, automatically or semi-automatically, and discover repetitive patterns, trends or rules that explain the behavior of the text.

2.3.1. Text mining in previous literature

In view of the modern consumer product environment, with a high degree of product assortment and large amounts of available data, France and Ghose (2016) highlight the need for data-driven tools to aid the analysis of brand competition. A small number of papers explore brand positioning using data-driven tools to analyze the text of online reviews.

Heng et al. (2018) studied review helpfulness incorporating a LDA analysis to extract topics from coffee reviews on Amazon.com. The authors use the extracted topics, together with traditional numeric review variables (e.g. rating and length), as independent variables in regression analyses to study review helpfulness. Nasiri and Shokouhyar (2021) collected online reviews from two US e-commerce websites to compare consumers' opinions with respect to refurbished versus brand new smartphones, analyze the customer satisfaction dimensions of most concern to consumers, analyze consumer perceptions of refurbished smartphones through opinion mining, and analyze the perceived benefits and risks of purchasing a refurbished smartphone by mining sentiment words. The authors also used latent dirichlet analysis (LDA) in their text mining analysis.

The subject of brand image and brand positioning has been addressed in several papers, the majority of which have adopted machine learning techniques for text mining. Guo et al. (2017) studied the positioning of hotels using Correspondence Analysis (CA) CA?? and attributes extracted from online reviews using LDA. Liu et al. (2017) also used LDA to extract the main attributes associated with brands in various categories (fast food, department store, footwear, telecommunications, and electronics) by analyzing and comparing four brands from each category. Wang et al. (2018) also conducted a LDA to compare the associated attributes of two competing wireless mouse products. Gensler et al. (2015) studied the McDonald brand image using the machine learning method of natural language processing (NLP), known as tokenization. Ahani et al. (2019) used a machine learning method, the Order of Preference by Similarity to Ideal Solution (TOPSIS) Technique adapted for text mining to study tourist satisfaction in the Canary Islands. Wong and Qi (2017) studied destination image in the tourism sector, applied to the case of Macau, using two specific software packages, *NVivo10* and *IBM_ManyEyes* for text mining. Kim and Kang (2018) used two machine learning algorithms, Latent Semantic Analysis (LSA), which is a supervised algorithm, and LDA, which is an unsupervised algorithm, to study the attributes associated with Korean and non-Korean cosmetics products. Others, such as Moon and Kamakura (2017), have analysed the positioning of a set of hotels in Manhattan using the main topics extracted from online reviews using LDA. Moon et al. (2021) conducted a segmentation of both reviewers and businesses based on online reviews using LDA to extract the main topics.

Table 1 gives a summary of some of the most common text mining tools used in previous literature sorted by research objectives, and adapted from Berger et al. (2020). Following this line, a few papers have conducted research to bring together disjoint bodies of literature on unstructured data analysis and text mining in social media (Balducci and Marinova, 2018; Berger et al., 2020; Hartmann et al., 2019; Kübler et al., 2019). Balducci and Marinova (2018) contextualized the different types of unstructured data, such as images, text, video and voice, providing a synthesis of the characteristics of each type of data and its use in different areas of marketing. Other scholars, such as Berger et al. (2020), focus on a specific type of unstructured data, i.e., text, and provide a review and discussion on the different methodologies used in text mining analysis.

Table 1
The commonest text mining approaches and tools used in the literature, by research objectives (Berger et al., 2020).

Main Objective	Specific research objectives	Type of text mining method	Text mining tools	Papers
Entity (Word) extraction Extracting and identifying single words	Sentiment analysis Identifying psychological associations Consumer and market trends	Lexicon-based methods Machine learning algorithms	Dictionaries and lexicons (e.g., LIWC, NRC Emotion Lexicon, BING, AFINN, SentiWordNet) Machine learning classification tools (e.g., deep learning)	Lee and Bradlow (2011) Ludwig et al. (2013) Kübler et al. (2019) Mahr et al. (2019) Zhang (2019)
Topic extraction Extracting the main topics discussed in the text	Summarizing the discussion Perceived product features Consumer needs and market trends	Machine learning algorithms	Machine learning algorithms: LDA (Latent Dirichlet Allocation) LSA (Latent Semantic Analysis)	Chen et al. (2015) Guo et al. (2017) Heng et al. (2018) Puranam et al. (2017) Tirunillai and Tellis (2014) Wang et al. (2018)
Relation extraction Extracting and identifying relationships between words	Identifying problems mentioned with specific product features Identifying product attributes mentioned positively/ negatively Identifying events and consequences.	Machine learning algorithms	Supervised machine learning Deep learning	Schnittka et al. (2012) Gensler et al. (2015) Netzer et al. (2012)

It is a general observation in the literature that papers applying text mining techniques fall into two categories; those using lexicon-based approaches and those using machine learning algorithms. Hartmann et al. (2019) and Kübler et al. (2019) compared the performance of several text classification methods, including lexicon-based approaches (e.g. LIWC by Pennebaker et al., 2007) and machine learning algorithms (e.g., random forest and naïve Bayes). Hartmann et al. (2019) concluded that machine learning algorithms tend to perform somewhat better than lexicon-based approaches, but that the difference in accuracy is often only slight. Thus, the selection of one type of approach or another necessarily depends on the research objectives and available resources, always bearing in mind that lexicon-based methods are quicker and easier to apply. Managers, in many cases, are faced with a tradeoff between cost and a deeper understanding of social media content (Kübler et al., 2019).

The general observation is that most previous scholars have applied the LDA technique when studying the content of online reviews, thus

revealing their focus on revealed product attributes. However, the study of more psychological perceptions of products and brands is less common. It is also noticed that past studies have largely limited their analysis to a small number of *a priori* selected products or brands. Our research is designed to deal with several products from several brands. Moreover, since the literature is quite heterogeneous in terms of text-mining processes and techniques, it falls short in providing clear guidelines for prospective scholars and practitioners in the field. We attempt to overcome this issue by suggesting a structured procedure for research into the text content of online reviews within the context of brand image and positioning.

3. Research procedure

The main objective of this research is to present and illustrate a unified and structured procedure for drawing relevant conclusions on the subject of brand image and positioning from the text of online reviews. The proposed procedure could serve as an easy-to-follow guide for small and medium firms, which, while lacking the necessary knowledge and expertise to approach text mining with a more sophisticated machine learning tool, nevertheless wish to benefit from the embedded content in their available eWOM data. The stages in the proposed procedure, together with a summary of the tasks involved in each stage, are shown in Table 2.

In the following sub-sections we describe the methodology applied in each stage of the procedure proposed in this study, and summarized in the flowchart in Fig. 1. A detailed description of the steps followed and the R software options selected is included in Appendix A, under the heading "*Brand positioning from online reviews: steps and code*".

3.1. Data acquisition

For the purposes of our research, we collected a total of 62,496 online consumer reviews from the website of a US cosmetics retailer, ranked among the top-50 shopping sites in the US in March 2017 according to *alexa. com*. The data were obtained by means of Web Scraping, a data mining technique for the automated collection of structured web data also known as Web data extraction, screen scraping or Web harvesting. A scraping tool typically makes HTTP requests to a target website and extracts the data from a page. It typically parses content that is publicly accessible and visible to users and rendered by the server as HTML. The Web Scraping process involves the following steps: identify the target website (www.sephora.com), collect URLs of the pages from which data is to be extracted (product webpages), make a request to these URLs to get the HTML of the page, use locators to find the data in the HTML (e.g., find reviews for each product) and save the data in a CSV file.

The collected database represents the entire set of online reviews for the blusher category available at the website on February the 17th 2017. These reviews refer to a total of 131 products from 44 different cosmetics brands.

Fig. 2 shows a typical online consumer review on the online cosmetics retailer's website, where both textual and non-textual cues can be observed. Review text is marked in blue in Fig. 2 since it is the part of the review on which we are focusing in this research.

As well as the text from individual online reviews, we collected other non-textual information to describe the sample used in this research: brand name, average price, position in the website's bestseller list,[1] average brand rating, number of brand reviews and average length of brand reviews. Having obtained these descriptive data, it is interesting to contextualize the relevance of the brands in our sample. To account for brand popularity and to form an idea of consumers' perceptions in this respect, data were gathered to estimate the number of followers of each brand on Instagram (Social Blade, 2017). Instagram was chosen because every brand in the category was featured on it and because it is the fastest growing social network site globally and one of the most used both by companies and users to share beauty-related photographs and videos (Sheldon and Bryant, 2016; SproutSocial, 2018). However, in the absence of a brand's Instagram presence, other social networks, such as Twitter or Facebook, could be used to measure brand popularity. Otherwise, sales or market share information would have to suffice for the purpose. Table 3 shows the brands represented in the blusher category on the online retailer's website together with some non-textual brand information.

Table 2
Stages and tasks involved in the proposed research procedure.

Stage	Tasks involved in each stage
1. Data acquisition	- Downloading online reviews and other relevant information from the website of interest. Several options are available to build the online review database (if online reviews are on the retailer's own website, they are already available): - Web scraping the online reviews of interest. Tools such as the *rvest* package in R are available for scraping data from websites. - Some websites provide open-source databases of online reviews that are ready to download (e.g., Kaggle at https://www.kaggle.com/datasets). - Amazon Mechanical Turk (MTurk) is a crowdsourcing marketplace that makes it easier for private individuals and businesses to outsource processes and jobs to be performed virtually by a distributed workforce (Amazon Mturk at https://www.mturk.com/). Thus, web scraping tasks could be similarly outsourced. - If needed, other relevant information (e.g., industry reports, google trends and analytics websites) can be recorded from other websites.
2. Text Mining	● Text mining methods: - Machine learning algorithms (e.g., Latent Dirichlet Allocation). - Lexicon-based methods, such as LIWC (Pennebaker et al., 2015), WordStat (Peladeau, 2016) and NRC Emotion Lexicon (Mohammad and Turney, 2010).
3. Data aggregation	●Review variables should be aggregated when conducting either product-level or brand-level analyses. - *Issue to consider*: Deciding which products/brands should be included in the analysis. Some products/brands might have such a small number of online reviews, that the average might not be meaningful.
4. Brand Positioning: Building a Perceptual Map	● Dimensionality reduction techniques, such as Correspondence Analysis (CA), Multidimensional Scaling (MS) and Principal Component Analysis (PCA) are the most frequently used. ● Analyses can be conducted using R, a free statistics software package. R has specific packages to facilitate different types of analyses. Each package includes a user guide.
5. Brand Positioning: Identifying Brand Subgroups	● Different types of clustering methods are available (e.g., hierarchical clustering and k-means clustering). ● There are R software packages to suit each clustering method; each one comes with the necessary guidelines. ● Clustering results can be graphically displayed in a dendogram or perceptual positioning map. ● Clusters can be described by analysing variable means.

[1] The bestseller list offers a snapshot of sales through the online retailer for up to a week. A product's sales rank is inversely related to its sales, which means that the top product in the sales rank in a specific product category is the one with the highest sales during the previous week.

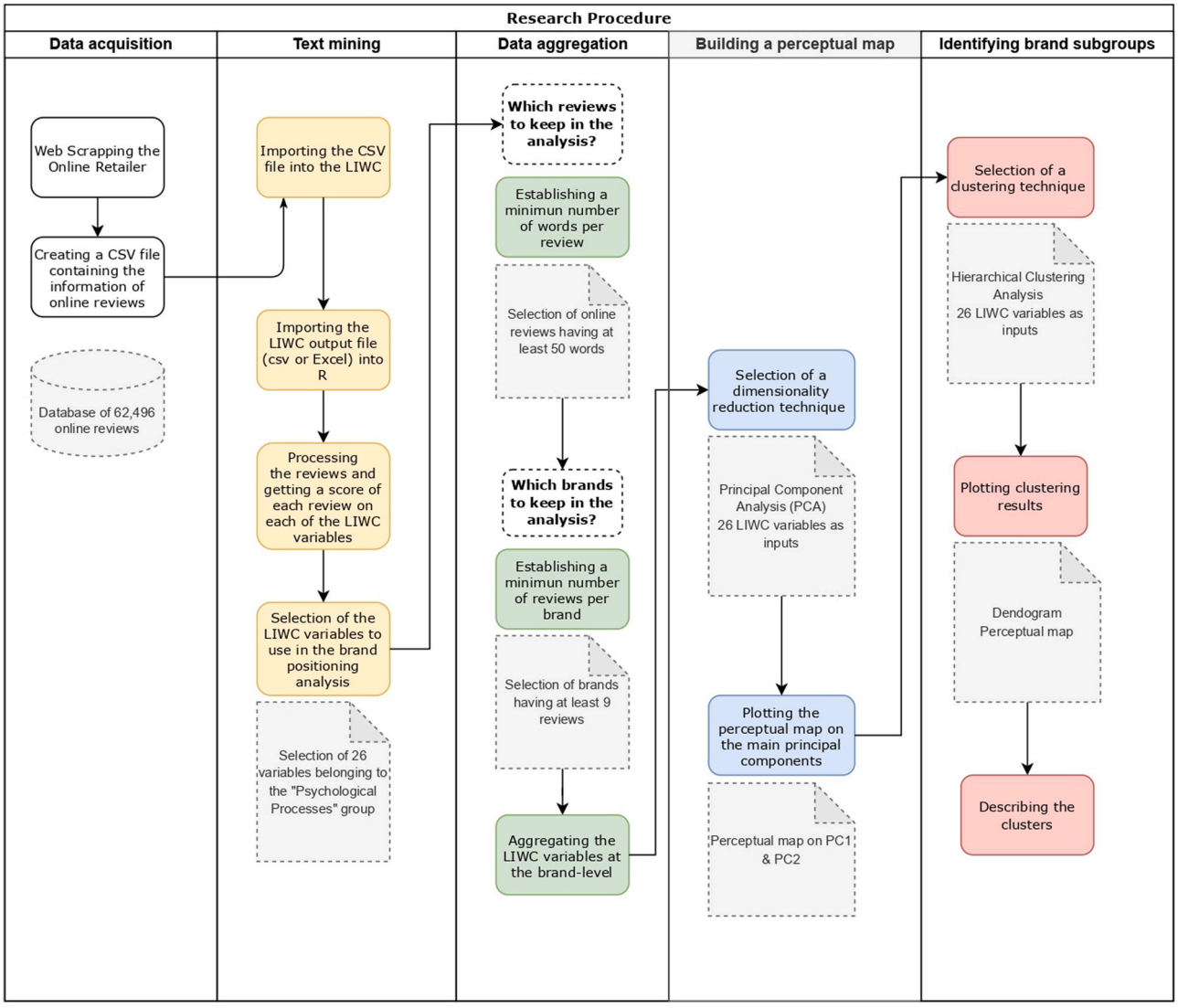


Fig. 1. Flowchart of the research procedure.

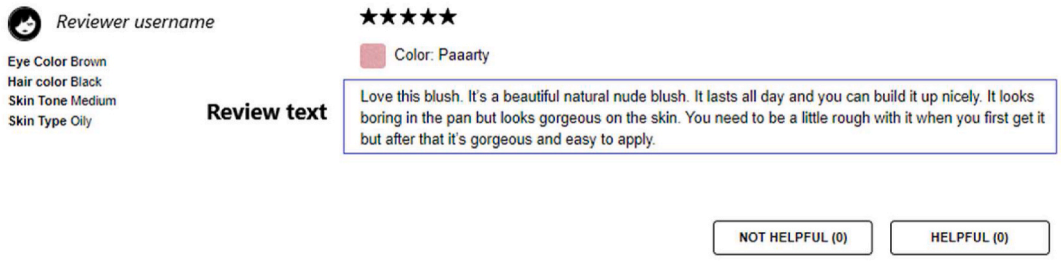


Fig. 2. Typical online review at the online retailer website.

3.2. Text mining: the use of LIWC

For the text mining analysis, this research adopts the lexicon-based Linguistic Inquiry and Word Count (LIWC) program developed by Pennebaker et al. (2007). The LIWC was initially devised to identify emotional content in the written answers of participants in a health survey (Pennebaker and Francis, 1996). Since the introduction of the first version, the LIWC has demonstrated proven category validity in hundreds of studies spanning dozens of psychological domains and has been widely used in the fields of Psychology and Marketing (Cohn et al., 2001; Ireland et al., 2011; Ludwig et al., 2013).

Of the approximately 90 data variables covered by the LIWC, some categories, such as articles and personal pronouns, are straightforward. However, other emotional and psychological dimensions are more

Table 3
Cosmetics brands included in the research and main descriptive statistics of non-textual variables.

Brand	Number of products associated with the brand	Average price (*euros* per *gram*)	Brand average rating	Brand's position in the online retailer's bestseller list	Total number of reviews of the brand	Average number of reviews per product associated with the brand	Average length of online reviews	Total number 0f brand's Instagram followers
bareMinerals	2	23.7	4.6	9	2051	1025.5	40.5	398,280
BECCA	3	5.1	4.5	17	1006	335.3	57.8	1,511,811
Benefit Cosmetics	11	3.2	4.4	19	9943	903.9	54.1	5,865,273
Bite Beauty	1	5.1	4.5	1	1547	1547.0	78.3	288,840
Black Up	1	7.8	4.7	42	21	21.0	46.3	56,769
Bobbi Brown	5	6.8	4.5	13	949	189.8	51.8	2,077,698
BURBERRY	3	7.2	4.6	26	109	36.3	76.4	8,698,164
Chosungah 22	3	2.2	3.3	38	32	10.7	83.5	11,333
Ciaté London	1	3.2	4.3	40	14	14.0	58.8	206,371
CLINIQUE	3	5.9	4.5	11	929	309.7	55.7	1,214,278
Dior	4	6.0	4.6	18	569	142.3	59.4	13,957,736
Estée Lauder	1	4.4	3.6	37	18	18.0	51.2	1,642,483
Giorgio Armani Beauty	1	2.4	3.7	14	89	89.0	73.9	110,379
Givenchy	2	6.1	4.6	31	177	88.5	55.4	7,647,395
Guerlain	4	6.0	4.1	34	36	9.0	73.2	529,655
Hourglass	4	9.9	4.4	3	1800	450.0	65.7	760,788
ILIA	1	6.8	4.6	21	18	18.0	79.1	45,251
KEVYN AUCOIN	3	4.7	4.1	24	93	31.0	74.5	196,188
Lancôme	3	6.8	4.6	27	283	94.3	51.4	1,461,331
Laura Mercier	4	6.9	4.5	25	1356	339.0	46.3	1,134,996
MAKE UP FOR EVER	5	9.0	4.4	6	1644	328.8	56.5	3,382,912
Marc Jacobs Beauty	2	5.6	4.6	10	1635	817.5	56.9	3368
MILK MAKEUP	2	2.4	4.2	22	91	45.5	63.7	152,732
NARS	8	5.7	4.6	2	22043	2755.4	46.4	4,083,316
NUDESTIX	6	11.9	4.3	32	287	47.8	61.4	72,033
Perricone MD	1	4.1	4.5	44	238	238.0	49.9	35,715
Retailer brand	9	2.8	4.3	12	3058	339.8	74.3	10,474,573
rms beauty	1	5.2	4.8	7	4	4.0	45.1	100,986
Shiseido	1	5.3	4.1	43	8	8.0	10.34	198,447
Smashbox	4	4.5	4.1	28	1700	425.0	48.2	2,777,377
stila	3	5.7	4.4	29	1257	419.0	58.3	1,960,925
Supergoop!	1	4.9	4.5	35	10	10.0	65.2	17,486
surratt beauty	1	8.0	4.0	16	51	51.0	95.5	23,465
tarte	5	4.5	4.6	5	5138	1027.6	58.2	5,873,000
Tata Harper	1	8.4	4.1	23	41	41.0	71.3	83,349
The Estée Edit	1	4.7	4.1	36	33	33.0	48.8	1,642,483
TOM FORD	2	12.2	4.2	33	39	19.5	88.9	8468
Too Cool For School	3	1.9	3.8	41	35	11.7	80.7	10,836
Too Faced	4	4.8	4.2	4	1658	414.5	54.1	8,210,500
trèStiQue	2	5.6	4.7	39	3	1.5	104.3	15,935
Urban Decay	3	2.7	4.3	8	2127	709.0	56.9	7,690,863
Viseart	1	3.3	5.0	30	9	9.0	110.9	72,634
Wander Beauty	2	7.1	4.1	20	65	32.5	73.6	33,929
Yves Saint Laurent	2	4.2	4.4	15	282	141.0	66.8	2,246,873

(1) Brand average rating out of 5; Number of Instagram followers on February 17th, 2017.

subjective.[2] The software has been used and the validity of the methodology confirmed in over 100 studies analysing online content such as instant messaging (Ireland et al., 2011; Ludwig et al., 2013) and online blogs (Cohn et al., 2001). The program contains a dictionary of approximately 4500 words covering a number of dimensions. By means of the word count strategy, texts are analysed on a word-by-word basis, each word being compared against the pre-dictionary. The linguistic indicator scores for each LIWC variable are calculated as the percentage of words that match the pre-defined dictionary. To measure the degree of positive emotions in an online review, for example, the LIWC calculates the total number of times the words defined in the dictionary as pertaining to positive emotions (e.g., "love", "nice" and "beautiful") appear in the review and the result is divided by the total number of words in the text. For further explanation, we include below a couple of examples extracted from our database. The variable *Body* measures the mentions made in the review to body-related issues. For example, the score on the *Body* variable in the following review is higher in

[2] In those cases, human judges were required to evaluate the words suitable for each category. For subjective categories, an initial set of word candidates for each category was built from dictionaries, thesauruses, questionnaires and lists made by research assistants. See Tausczik and Pennebaker (2010) to get more information on how dictionaries are built. Then, groups of three judges independently rated if each candidate word was appropriate to each category. Finally, a word remained in the category if two out of the three judges agreed it should be included. A word was deleted from the category if at least two of three judges agreed it should not be included. The final agreement of the judges was 100%.

comparison to other reviews in our database:

> "I own both orgasm n super orgasm. If u like subtle shimmer, go with orgasm, because super orgasm contain serious gold glitter. When I wear the SO, I always tone down the eye n lips make up. So if u like to be heavy on eye or lips make up, better go with O".

The variable *Family* represents mentions made to family issues. For example, the online review shown below scores higher in the *Family* dimension than others in our database:

> "I bought several of these last Valentine's Day for me and my daughter and daughter-in-law and LOVE it. It is kind of a blush, a bit of a bronzer, but mainly just adds this luminescent glow of shimmer that is absolutely flattering. I liked it so much that I just bought them as birthday gifts for my niece and sister-in-law"

Although several lexicon-based tools are available (e.g. SentiWordNet, AFFIN, NRC Emotion Lexicon, etc.), we propose to use the LIWC for several reasons.

First of all, it is easy to implement (Hartmann et al., 2019) and the license is freely available to anyone. Moreover, the software for handling the texts is very intuitive. The user only needs to import the texts from any file type (e.g., word, pdf, excel, etc.) and the LIWC automatically analyzes them and assigns a score to each text and each output variable. Thus, the LIWC does not require any knowledge of machine learning methods. The LIWC is therefore a powerful tool to support non-technical users when conducting the text mining tasks involved in our proposed research procedure.

In addition, it informs on some 90 categories of text variables. Thus, it has the capacity to extract a vast quantity of insights from the content of online reviews. Other lexicon-based available tools, such as AFINN and SentiWordNet, only classify words into two categories: positive or negative. The National Research Council (NRC) lexicon goes a bit deeper and categorizes words into more emotional categories, beginning with two general sentiment categories: *positive* and *negative*, while also adding eight emotions: *anger*, *anticipation*, *disgust*, *fear*, *joy*, *sadness*, *surprise,* and *trust*, thereby increasing the specificity of the analysis. Even so, these other tools do not cover as many variables as the LIWC, which includes aspects such general descriptors (e.g. words per sentence), standard linguistic dimensions (e.g., pronouns, articles, adverbs and auxiliary verbs) and information pertaining to the so-called "Psychological Processes" group. As claimed in the related literature, the term "brand image" covers aspects relating not only to physical and functional attributes but also to the emotional and psychological benefits of the brand (Delgado-Ballester and Fernández-Sabiote, 2016; Ruane and Wallace, 2015; Simon et al., 2016). In this regard, the LIWC performs well by providing information on 53 variables under the "Psychological Processes" category.

Moreover, Hartmann et al. (2019) compared the performance of the LIWC to other more sophisticated machine learning algorithms for text mining and concluded that the LIWC performs practically as well as machine learning algorithms, which require higher computational skills.

3.2.1. Selection of LIWC variables

LIWC provides a total of approximately 90 output variables. Some of them are general descriptors (e.g., words per sentence), others are standard linguistic dimensions (e.g., pronouns, articles, adverbs, prepositions and auxiliary verbs) and 53 variables belong to the so called "Psychological Processes" group. Within this group of 53 variables, there are 10 general variables or categories and 43 more specific variables or sub-categories. In this research, we use a set of 26 variables, which belong to 7 general variables or categories under the "Psychological Processes" group: *affect*, *social processes*, *perceptual processes*, *biological processes*, *drives*, *relativity* and *personal concerns*. General descriptors and standard linguistic dimensions were not included in this research, since they do not contribute to brand image information. Some LIWC variables from the "Psychological Processes" group have also been excluded from this analysis because they measure issues relating to narrative style rather than brand image, which has been said to include physical attributes and functional, emotional and self-expressive benefits. These are represented by the following general and specific variables: *cognitive processes,* which include the specific variables *insight*, *causation*, *discrepancy*, *tentative*, *certainty* and *differentiation*; *informal language,* which includes the specific variables *swear words*, *netspeak*, *assent*, *nonfluencies* and *fillers;* and *time orientation*, which includes the specific variables *past focus, present focus* and *future focus*. The subcategory of *death*, which belongs to the category of *personal concerns*, does not feature in our research because it is irrelevant in our study context. Appendix B shows the complete list of 53 variables comprising the LIWC Psychological Processes category.

The overall category of *affect* in the LIWC has two subcategories: *PosEmotions* and *NegEmotions*. *NegEmotions,* in turn, is divided into three groups: *anger*, *anxiety* and *sadness*, whereas *PosEmotions* is broken down no further. Thus, to ensure a balanced array of sentiments, we only use the general categories of *PosEmotions* and *NegEmotions* as variables in our empirical analysis.

The excluded categories could, nevertheless, be included in future studies if they align with the specific research objectives. For example, if the interest lies in the communication style of the reviewer (e.g. formal, informal, tentative, certain or use of swear words) it might be worth exploring the variables in the *informal language* category, and to discriminate between references to past experiences, current situations and future expectations the *time orientation* category could be useful.

3.3. Data aggregation

The LIWC was used to analyze the text of every online review in our database, which comprises a total of 62,496 online reviews about 44 different brands. Nevertheless, since most word counting methods deal with percentages, experts in the field of psychological text analysis generally recommend a minimum number of words per text. Indeed, one of the founders of the LIWC recommends a minimum word count of 25–50 words per text (Boyd, 2017). Other authors, such as Webster et al. (2019), consider a cut-off threshold of about 100 words to be reasonable for most studies, although a lower number might sometimes be justified, and there is no minimum cut-off rule. According to Boyd (2017), one case for lowering the cut off point to below 25–50 is when dealing with social media posts or online opinions, which tend to be short. On the other hand, if working with texts such as online articles or blog entries, which are generally longer, the cut-off level could be raised above 50 words. In this research, where the average review length is 52 words, and based on previous recommendations, we set the cut-off point at 50 words,[3] thus including all reviews around and above the average length. Nevertheless, other cut-off values are possible, depending on the research objective or text type (e.g., lower cut-off values might be set when dealing with tweets, and higher values could be used when dealing with blog entries). Having set a cut-off value of 50 words, our final database comprises a total of 25,549 online reviews for 44 brands.

To explore brand positioning, we worked with brand average variables, therefore the LIWC output, which consists of a score for each review, was aggregated into a brand level, as per the following equation:

[3] We also ran the analysis on the entire database (with no cut-off value), finding that the descriptive statistics of the LIWC brand variables remained virtually unchanged. The impact on the results, therefore, is probably negligible Thus, in the specific context of blushers, brand associations (in terms of LIWC variables) vary little with the number of words in the review. With this in mind, the final analysis was conducted on reviews with more than 50 words to increase statistical accuracy. More words mean more reliability and accuracy when quantifying psychological processes, which is difficult with any research method (Boyd, 2017).

Table 4
Text variables extracted from LIWC and brand average descriptive statistics of text scores.

Variable	Example of words	N	Mean	Sd	Min	Max
Affect						
PosEmotions	Amazing, benefit, excellent, fair	41	5.642	0.991	3.842	9.094
NegEmotions	Afraid, anxious, cruel, despair	41	0.658	0.207	0.000	1.049
Social processes						
Family	Cousin, honeymoon, marry, husband	41	0.053	0.065	0.000	0.386
Friend	Beloved, best friend, bud, classmate	41	0.064	0.081	0.000	0.470
Female	Bride, daughter, ex-wife, girl	41	0.145	0.107	0.000	0.467
Male	Boy, brother, fellow, man	41	0.023	0.021	0.000	0.081
Perceptual processes						
See	Appear, beauty, colour, shine	41	4.532	0.753	2.576	6.445
Hear	Listen, noise, quiet, speak	41	0.158	0.064	0.000	0.302
Feel	Cold, dry, hard, hot	41	2.036	0.421	1.573	3.696
Biological processes						
Body	Cheek, eye, face, facial	41	2.560	1.037	0.683	5.519
Health	Acne, allergy, pain, fitness	41	0.158	1.104	0.000	0.642
Sexual	Lover, nude, sexy	41	0.115	0.167	0.000	0.794
Ingestion	Diet, eat, fat, food	41	0.234	0.185	0.000	0.841
Drives						
Affiliation	Belong, colleague, reunion, party	41	0.931	0.389	0.116	2.463
Achievement	Able, ambition, confident, proud	41	1.347	0.289	0.781	2.326
Power	Beat, celebrity, comply, win	41	1.282	0.328	0.357	2.103
Reward	Achieve, advantage, benefit, earn	41	1.793	0.354	0.481	2.725
Risk	Alarm, avoid, dangerous, doubt	41	0.227	0.105	0.000	0.624
Relativity						
Motion	Approach, arrive, attend, carry	41	1.329	0.304	0.510	1.931
Space	Anywhere, back, big, broad	41	5.856	0.719	3.588	7.495
Time	After, age, always, never	41	3.840	0.761	1.933	5.742
Personal Concerns						
Work	Business, class, company, student	41	1.569	0.361	0.880	2.397
Leisure	Art, band, café, party	41	0.277	0.132	0.000	0.754
Home	Family, home, house, neighbour	41	0.062	0.058	0.000	0.258
Money	Affordable, bargain, buy, cheap	41	1.163	0.296	0.656	1.966
Religion	God, bless, demonic, karma	41	0.033	0.036	0.000	0.164

$$T_b = \frac{\sum T_n}{N_b} \qquad \text{Equation 1}$$

where *Tb* denotes the average text score for brand *b*; *Tn* is the text score for review *n* for brand *b* and N_b denotes the total number of online reviews for brand *b*.

Brand average values are used to study brand image and positioning patterns. Initially, the LIWC assigns a score to each of its variables in every individual consumer review. However, we are interested in knowing not how each individual consumer perceives a brand, but how the brand is perceived in the overall market. We therefore aggregated the data to brand level. Some scholars, such as Li and Hitt (2008) deal with this issue by excluding any observations where the average is based on fewer than three reviews. In a similar line, Moon et al. (2021) consider only those cases with more than ten reviews. The rationale behind these decisions is that a minimum number of online reviews is needed to determine a meaningful trend. However, the literature offers no rule of thumb for setting the minimum. The decision might depend on factors such as the characteristics of the reviews in the database, the average number of reviews per product and the specific research objectives. All this considered, it is worth noting that the more data used to determine the pattern, the higher the validity of the results (Boyd, 2017). In our case, we decided to exclude any brands whose products average fewer than nine reviews ($N < 9$). This threshold corresponds to brands at the fifth percentile in terms of the average number of product reviews, which is 8.15 reviews. Thus, three brands: trèStiQue, rms beauty and Shiseido were removed from the study sample, and the number of brands for analysis decreased from 44 to 41.

3.4. Brand positioning: building a perceptual map

When building perceptual positioning maps literature has usually adopted dimensionality reduction techniques. In both Statistics and Machine Learning, the number of input variables in a dataset is known as its dimensionality. Dimensionality reduction methods are classified into those used to retain only the most important features (backward elimination, forward selection and random forest), and those used to find a combination of new features, which can, in turn, also be divided into linear (e.g. Principal Component Analysis and Factor Analysis) and non-linear techniques (e.g. Kernel PCA and t-SNE).

In this research, we propose to use a linear method, Principal Component Analysis (PCA), to linearly project the original data onto a low-dimensional space. Linear dimensionality reduction can be used to visualize and explore structure in the data and extract meaningful feature spaces (Cunningham and Ghahramani, 2015; Gwin and Gwin, 2003). Brand mapping consists of graphing the position of competing brands in a market on the basis of their location space, as defined by the key dimensions (Tirunillai and Tellis, 2014). Overall, PCA is a technique that transforms a set of correlated variables (p) into a smaller k ($k < p$) number of uncorrelated variables called principal components, while retaining as much of the variation in the original dataset as possible.

In this research, the PCA is run in R software using the "*factoextra*" package (Kassambara and Mundt, 2020). The 26 variables (p) resulting from the LIWC are used as input variables to derive the brand perceptual positioning map using Principal Component Analysis (PCA).

3.5. Brand positioning: identifying brand subgroups

Although PCA enables us to see similarities between brands based on the different principal components, a clustering analysis is required to clearly identify brand subgroups in the blusher category. The object of clustering is to create clusters of brands such that there is as much similarity as possible within each cluster and as much dissimilarity as possible between different clusters (Kaufman and Rousseeauw, 2009; Maimon, 2005).

In this research, given its popularity, we adopted an agglomerative hierarchical clustering technique, using Wald's clustering algorithm. Wald's method aims to minimise the total within-cluster variance (Hair et al., 2010). One of the main advantages of hierarchical clustering over non-hierarchical clustering is that it is easy to understand and implement. The hierarchical clustering algorithm results in a dendrogram that can be used to understand the big picture as well as the groups within the data. By evaluating the treelike structure of the dendogram, researchers can easily evaluate any of the possible clustering solutions from one analysis (Hair et al., 2010).

The variables used as inputs in the hierarchical clustering algorithm are the 26 text variables previously used in the PCA, which represent psychological brand associations. Before running the hierarchical clustering algorithm, each text variable was scaled. As noted by Hair et al. (2010), standardization is important when variables are measured in different scales but it can be also used to facilitate the interpretation of

Table 5
Correlation matrix for the text variables.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1.PosEmotions	1																									
2.NegEmotions	−0,46*	1,00																								
3.Family	0,50*	0,05	1,00																							
4.Friend	−0,21	0,12	−0,23	1,00																						
5.Female	0,14	−0,30*	0,34*	0,26	1,00																					
6.Male	0,09	0,17	0,00	−0,13	−0,17	1,00																				
7.See	0,04	−0,42	−0,11	−0,25	0,35*	−0,06*	1,00																			
8.Hear	−0,16	−0,08	−0,06	0,19	0,04	−0,25	0,21	1,00																		
9.Feel	0,11	−0,32*	−0,09	−0,17	0,00	−0,21	0,42*	0,12	1,00																	
10.Body	0,52*	−0,36*	0,36*	−0,18	0,03	−0,17	−0,16	−0,13	−0,09	1,00																
11.Health	0,46*	−0,05	0,63*	0,12	0,28*	−0,08	−0,33*	−0,19	−0,08	0,38*	1,00															
12.Sexual	−0,03	0,15	−0,01	0,04	−0,11	−0,05	−0,16	0,21	−0,21	0,25	−0,11	1,00														
13.Ingest	0,31*	−0,03	0,17	−0,22	−0,18	0,02	−0,22	−0,40*	0,04	0,38*	0,44*	−0,05	1,00													
14.Affiliation	0,78*	−0,20	0,66*	−0,36*	0,20	0,22	−0,07*	−0,14	−0,18	0,42*	0,38*	0,03	0,19	1,00												
15.Achievement	−0,23	0,04	−0,24	0,33*	−0,13	−0,09	−0,27	−0,08	−0,24	0,03	0,20	−0,16	0,07	−0,25	1,00											
16.Power	−0,43*	0,22	−0,44*	−0,10	−0,37*	0,17	0,07	−0,02	0,42*	−0,56*	−0,45*	−0,14	−0,26	−0,50*	−0,15	1,00										
17.Reward	0,54*	0,00	0,46*	−0,11	0,22	0,19	−0,20	−0,46*	−0,29*	0,36*	0,50*	−0,01	0,30*	0,58*	0,04	−0,48*	1,00									
18.Risk	−0,21	0,54*	0,17	0,51*	0,02	−0,02	−0,52*	−0,12	−0,45*	−0,16	0,40*	−0,03	−0,09	−0,09	0,45*	−0,09	0,13	1,00								
19.Motion	0,11	−0,03	0,18	−0,18	0,00	−0,07	−0,33*	−0,27*	−0,42*	0,22	0,09	0,10	0,13	0,46*	0,18	−0,48*	0,40*	0,15	1,00							
20.Space	−0,34*	0,32*	−0,23	−0,49*	−0,39*	0,14	−0,11	−0,28	0,01	−0,07	−0,23	0,11	0,20	−0,10	−0,16	0,45*	−0,15	−0,14	0,16	1,00						
21.Time	−0,26*	0,26	−0,26*	0,41*	0,14*	0,06	−0,07	−0,09	−0,31*	−0,12	−0,05	0,12	−0,14	−0,14	0,38*	−0,16	0,11	0,27*	0,14	0,00	1,00					
22.Work	−0,14	0,15	−0,16	−0,05	−0,36	−0,06	−0,35*	−0,27*	0,08	0,32*	0,03	0,06	0,34*	−0,11	0,22	0,05	−0,12	−0,07	0,07	0,39*	0,03	1,00				
23.Leisure	0,31*	0,14	0,17	0,31*	−0,09*	−0,26*	−0,49*	−0,18	−0,06	0,12	0,46*	0,06	0,49*	0,14	0,17	−0,27*	0,35*	0,37*	0,18	−0,26*	0,01	0,11	1,00			
24.Home	0,50*	−0,21	0,45*	−0,12	0,37	0,05	0,13	−0,22	0,03	0,16	0,36*	−0,16	0,11	0,62*	−0,24	−0,32*	0,44*	0,07	0,26*	−0,03	−0,28*	−0,17	0,14	1,00		
25.Money	−0,18	−0,30*	−0,18	0,08	0,05	−0,15	0,19	0,43*	−0,06	−0,28*	−0,27*	0,06	−0,43*	−0,21	−0,15	0,13	−0,36*	−0,14	−0,21	−0,22	−0,16	−0,43*	−0,23	−0,21	1,00	
26.Religion	−0,12	−0,05	−0,12	−0,15	−0,06	0,05	0,07	0,23	−0,28*	0,02	−0,10	0,38*	−0,27*	−0,01	0,27*	−0,12	0,13	−0,06	0,19	0,03	0,21	−0,24	−0,27*	−0,20	0,28*	1,00

the output even if the scales are the same. In our case, following the recommendation of Hair et al. (2010), we decided to standardize the variables having observed differences in the means and standard deviations of the clustering variables. Moreover, since our text variables are continuous, and because it is the most commonly recognized measure of similarity, we adopted the Euclidean distance to calculate the similarity between clusters (Hair et al., 2010).

An important decision when running the algorithm concerns the choice of method for determining the optimal number of clusters (k) in the data set. For this, we chose the Elbow method, which sees the percentage of variance explained as a function of the number of clusters. Thus, the total within-cluster sum of squares (WSS) is a function of the number of clusters, such that the optimal number of clusters is reached when the addition of one more does not greatly improve the total WSS (Bholowalia and Kumar, 2014). For our database, the output of the Elbow method indicates that the statistically optimal number of clusters is k = 4.

However, as claimed by Hair et al. (2010), no standard objective selection procedure exists and "*the selection of the final cluster solution requires substantial researcher judgment and is considered by many as too subjective. Even though sophisticated methods have been developed to assist in evaluating the cluster solutions, it still falls to the researcher to make the final decision as to the number of clusters to accept as the final solution*" (Hair et al., 2010). Therefore, although we might have a statistically optimal number of clusters, the decision as to how many clusters to retain might be based on other subjective decisions, such as the specific research objectives, the type of market or the type of variable explored.

4. Findings

4.1. Descriptive statistics of the text variables

Table 4 reports the descriptive statistics of the 26 text variables used in this research, together with examples of representative words from each category and descriptive statistics, once LIWC review scores were aggregated to brand averages. *N* represents the number of brands analysed. As they represent brand averages, the minimum statistic belongs to the brand with the minimum average score on each text variable and the maximum statistic belongs to the brand with the highest average score on each feature. For example, the variable *PosEmotions* measures the degree of positive emotion associated with each specific brand, a high value in this category means that consumers associate the brand with positive experiences; the variable *Power* measures the number of power-related words that appear in the online reviews of each specific brand, and a high value in this category means that consumers associate the brand with power.

From Table 4, it can be observed that the online reviews in this category are highly associated with words representing *posEmotions*, which might indicate that the consumers who post online reviews are quite satisfied with the brand. Associations with *space* and *time* issues are also quite common. This might suggest that consumers make references to product usage experiences (where, when and process). Some perceptual (*see* and *feel*) and *body* associations are also quite relevant, which makes sense in the context of cosmetics consumption, where perceptual and body-related experiences are likely to be an important part of product usage. Associations relating to *affiliation*, *achievement*, *power* and *reward* are also quite relevant in our setting, which suggests that consumers experience feelings such as fulfilment or social recognition when using blusher products. In terms of personal concerns, consumers usually associate brands with *work* and *money* experiences.

4.2. Brand perceptual positioning map

4.2.1. Exploring correlations

PCA is usually used when the variables are correlated. Therefore, with our data, the first step is to analyze the correlations between the

Table 6
Relevance of the first 10 PCs.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Eigenvalue	2.33	1.93	1.73	1.51	1.38	1.23	1.07	1.04	0.96	0.92
Proportion of variance	0.21	0.14	0.12	0.09	0.07	0.06	0.04	0.04	0.04	0.03
Cumulative proportion of variance	0.21	0.35	0.47	0.55	0.63	0.69	0.73	0.77	0.81	0.84

Plotting a Perceptual Positioning Map.

Table 7
Text variable loadings on the 3-dimensions (PC1, PC2 and PC3).

	PC1 "Power and money"	PC2 "Perceptual processes"	PC3 "Social processes"
PosEmotions	−0.32	0.22	−0.02
NegEmotions	0.07	−0.34	−0.08
Family	−0.30	0.10	0.05
Friend	0.05	−0.23	0.36
Female	−0.12	0.17	0.32
Male	−0.01	−0.02	−0.14
See	0.11	0.38	0.06
Hear	0.16	0.14	0.27
Feel	0.10	0.27	−0.17
Body	−0.26	0.05	−0.08
Health	−0.31	−0.09	0.08
Sexual	0.00	−0.07	0.04
Ingest	−0.21	−0.08	−0.29
Affiliation	−0.34	0.15	−0.03
Achievement	0.00	−0.31	0.14
Power	0.30	0.01	−0.23
Reward	−0.33	−0.05	0.02
Risk	−0.07	−0.37	0.21
Motion	−0.20	−0.13	−0.02
Space	0.10	−0.09	−0.42
Time	0.04	−0.26	0.18
Work	−0.01	−0.20	−0.34
Leisure	−0.20	−0.21	0.03
Home	−0.26	0.16	0.00
Money	0.18	0.17	0.26
Religion	0.05	−0.02	0.17

text variables, shown in Table 5, where some high positive correlations (*Family/PosEmotions*, *Space/Work,* Body/*Power* and *Reward/Affiliation*, *NegEmotions/Risk* and *Family/Home*) can be observed. Therefore, the use of PCA is justified as a means to avoid potential multicollinearity problems.

4.2.2. Relevance of principal components

After the PCA, the next step is to look at the proportion of variance explained by each component. Table 6 shows the relative importance of the first 10 principal components (PCs) obtained from the PCA, which enables a dimensionality reduction from 26 to 10 variables, while retaining 84% of the variance in our data, 47% of which is already explained by just the first three components. Generally speaking, when applying PCA methods, researchers tend to pick the first two principal components to plot a two-dimensional positioning map. In our case, PC1 and PC2 account for more than a third (35%) of the variance in the data and, if we add PC3, the explained variance increases to almost half the total variance (47%).

The proportion of variance explained by each PC is one of the factors that determine the number of PCs retained for interpretation. However, other subjective factors, linked to the specific research objectives, might influence this decision. One such factor would be *a priori* selection of a number of factors of interest (Hair et al., 2010). There is no general rule of thumb for determining the required minimum of explained variance of retained PCs. Nevertheless, there are some general recommendations. Samuels (2016) for example, recommends a minimum of 50% of the explained variance. In this research, we keep the first three components for later interpretation. However, more might be retained in company or scholarly research, depending on the objectives.

Table 8
Brand loadings on PC1, PC2 and PC3.

	PC1 "Power and money"	PC2 "Perceptual processes"	PC3 "Social processes"	Cluster of the brand (k = 4)
bareMinerals	−0.84	0.71	0.12	1
BECCA	0.85	−0.07	1.46	1
Benefit Cosmetics	−0.89	−0.66	−0.12	1
Bite Beauty	−1.74	−0.94	−1.04	1
Black Up	2.32	5.56	3.68	2
Bobbi Brown	0.46	0.60	0.72	1
BURBERRY	0.67	1.06	−0.71	1
Chosungah 22	2.20	−2.95	−2.21	1
Ciaté London	−0.91	−7.20	5.15	3
CLINIQUE	0.46	0.33	0.11	1
Dior	0.20	1.04	1.23	1
Estée Lauder	3.28	4.52	−1.46	2
Giorgio Armani Beauty	1.83	−2.14	−2.12	1
Givenchy	1.01	1.61	0.85	1
Guerlain	1.52	−0.10	−0.27	1
Hourglass	1.39	0.70	0.85	1
ILIA	−3.22	−0.28	−2.48	1
KEVYN AUCOIN	1.36	0.19	−2.72	1
Lancôme	−0.74	0.70	0.33	1
Laura Mercier	0.65	0.67	1.00	1
MAKE UP FOR EVER	0.68	−0.34	0.36	1
Marc Jacobs Beauty	−0.78	0.71	0.01	1
MILK MAKEUP	−1.13	−0.39	−1.23	1
NARS	−0.05	0.01	1.87	1
NUDESTIX	−0.98	−1.12	0.57	1
Perricone MD	−0.10	−2.13	−0.94	1
SEPHORA COLLECTION	0.21	−0.21	0.46	1
Smashbox	−0.06	−0.63	−0.40	1
stila	−0.42	−0.10	−1.50	1
Supergoop!	−10.63	2.51	1.71	4
surratt beauty	2.45	−0.18	−0.57	1
tarte	0.00	−0.01	1.47	1
Tata Harper	−2.58	−0.62	−2.08	1
The Estée Edit	4.53	−2.12	0.65	1
TOM FORD	1.86	1.29	−2.34	2
Too Cool For School	−1.21	−0.47	−2.79	1
Too Faced	−1.01	0.18	0.29	1
Urban Decay	−0.29	−1.01	0.97	1
Viseart	2.22	1.23	3.29	1
Wander Beauty	−2.59	0.75	−1.49	1
Yves Saint Laurent	0.00	−0.70	−0.68	1

The main objective of the PCA is to obtain a positioning map, for which the PCA output provides the necessary information regarding the loadings (or values) of the text variables on the different dimensions. Table 7 reports the loadings of the text variables for the three principal dimensions: PC1, PC2 and PC3. In PCA, variable loadings are interpreted as the coefficients of the linear combination of the initial variables from which the principal components are constructed. The factor loadings aid interpretation of the relevance and impact of each variable in each principal component.

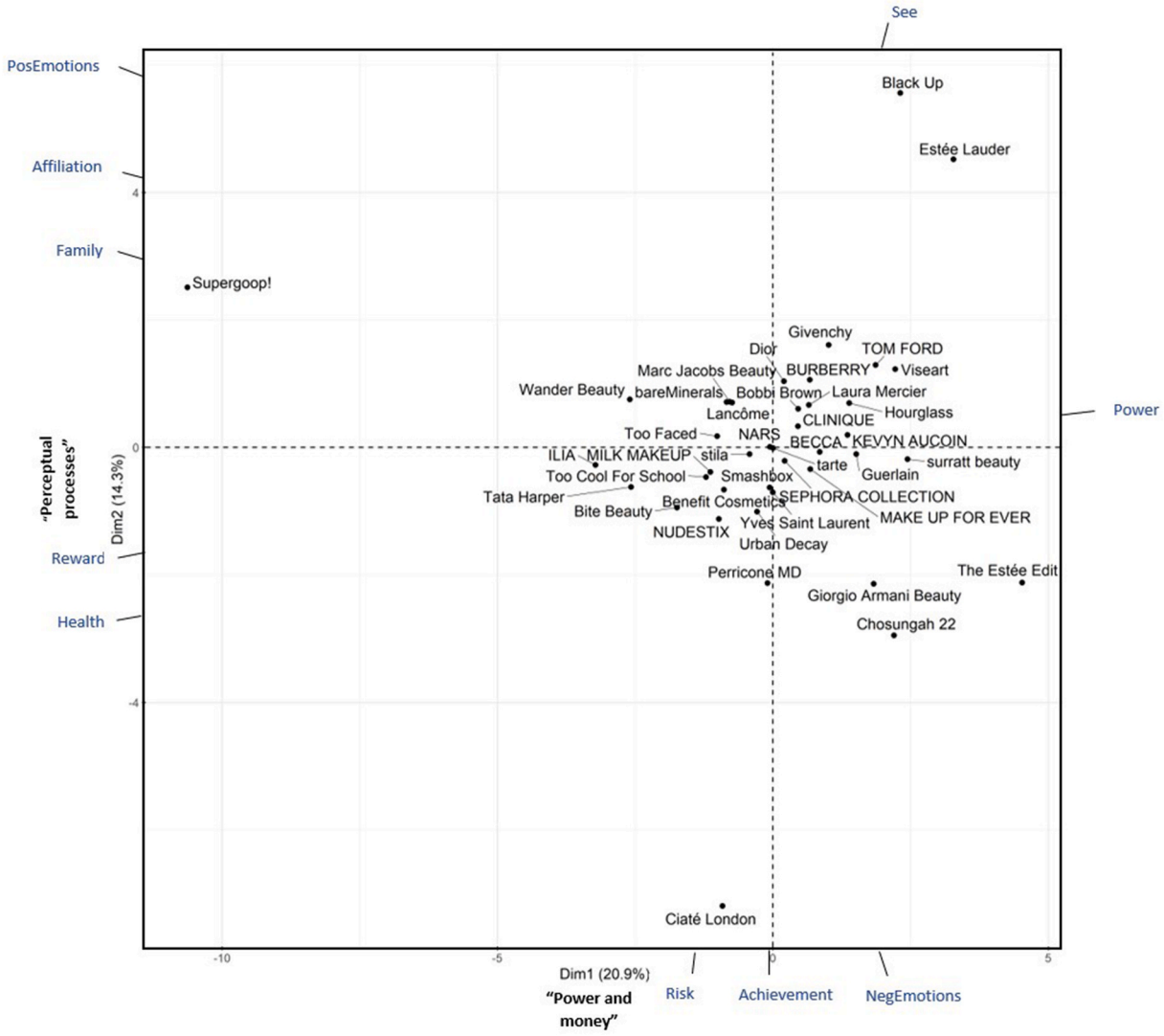


Fig. 3. Two-dimensional (2D) perceptual map on the two principal components (PC1 "Power and money" and PC2 "Perceptual processes"
Fig. 3. Two-dimensional brand positioning map on PC1 & PC2.

For interpretative purposes, the three main dimensions are given a label based on the text variables they represent. PC1 represents "*Power and money*" (high loadings for *Power* and *Money*); PC2 represents "*Perceptual processes*" (high loadings for variables such as *See* and *Feel*); and PC3 relates to "*Social processes*" (high loadings of variables such as *Friend* and *Female)*.

Table 8 shows the loading of each brand on the 3 principal components and the cluster in which each brand is included (based on the hierarchical clustering analysis performed in section 4.4). PCA loadings result either in a three-dimensional (3D) or two-dimensional (2D) perceptual or positioning map. For illustrative purposes and for ease of interpretation, we present and discuss the two-dimensional (2D) map on the two principal components, shown in Fig. 3. The main reason for not including a 3D perceptual map stems from the difficulty of displaying this type of map statically. Interpretation of the 3D plot might also prove somewhat difficult because there are many brands in close proximity to one another and the map cannot be rotated. For a more complete visualization, Appendix C presents the competitive landscape mapping based on PC1 & PC3 and PC2 & PC3. Again, more combinations of the various PCs might be interpreted in company or scholarly research, depending on the objectives.

The two-dimensional brand positioning map in Fig. 3 results from the combination of two inputs: first, the text variable loadings on PC1 (X-axis) and PC2 (Y-axis); and, second, the brand loadings on the two dimensions (PC1 and PC2). The textual factors represented at the edge of the perceptual positioning map are those that load heavily on PC1 and PC2. The text variables are plotted according to the loadings reported in Table 7, while the brands are plotted according to the loadings shown in Table 8.

From the brand positions on the map, we can examine their similarity in terms of the text variables. It can be seen, for example, that consumers strongly associate the brand BlackUp with *See* and *Feel*. The TOM FROD brand is highly associated with *Money* and *Hear* and the Perricone MD brand is positively related to *Work*. Note also that Supergoop! is positioned far from the other brands on the map and has strong associations with *Family*, *Affiliations, PosEmotions, Body* and *Home*. Its position indicates that this brand is highly differentiated in terms of the psychological brand associations with heavy loadings on PC1 and PC2.

Perceptual maps can also be used to identify competitors. In the case in hand, the close proximity of most of the brands might indicate that they produce similar psychological associations in consumers. For example, Dior is located close to Burberry. This indicates that customers perceive Dior and Burberry to be similar in terms of PC1 & PC2 and that the two brands compete strongly against each other in most associations. For a closer examination of the similarities between brands and to

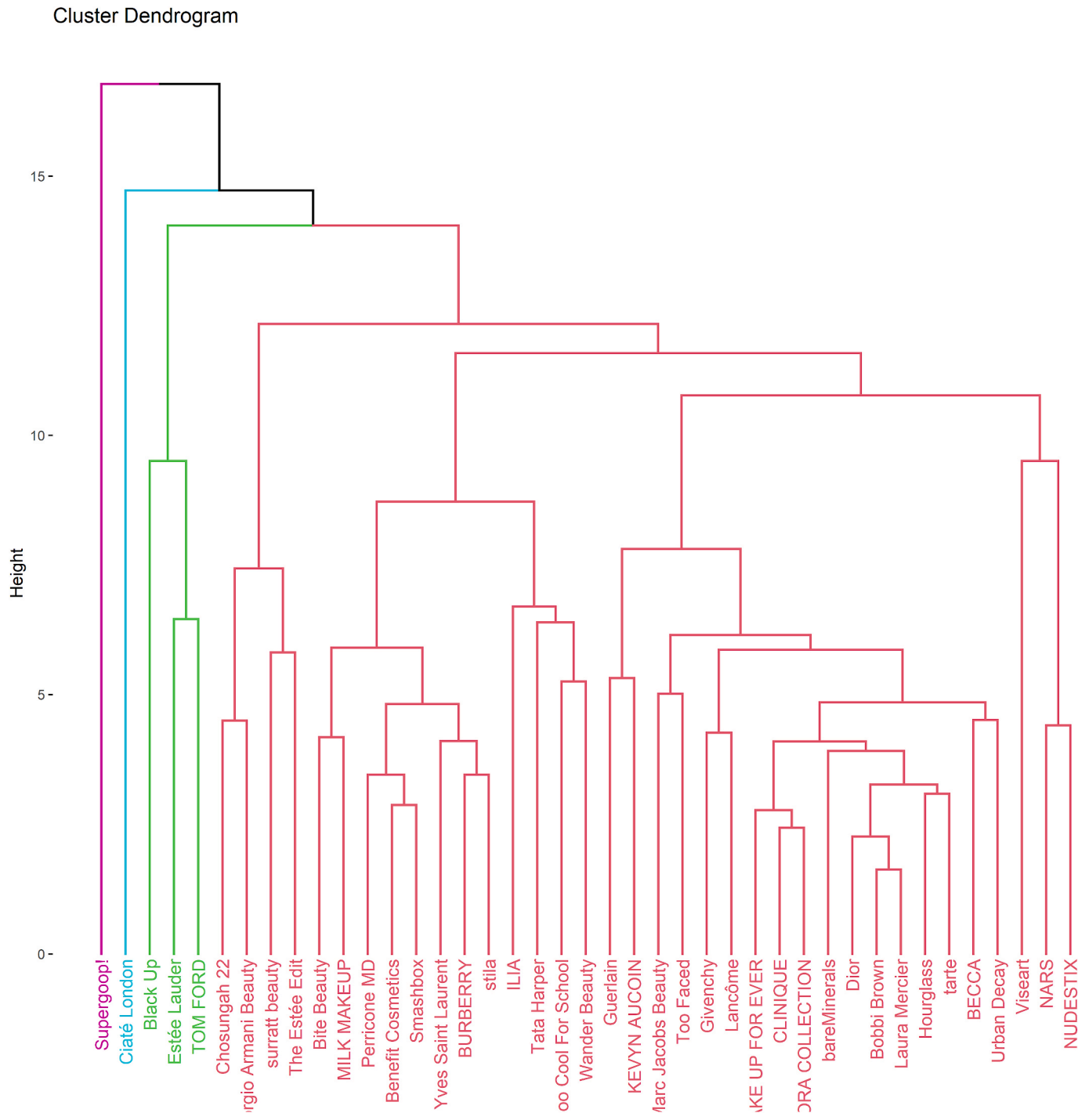


Fig. 4. Hierarchical clustering dendogram (Agglomerative algorithm: Linkage (Ward's method), k = 4).

identify possible brand subgroups, we carried out a clustering analysis.

4.3. Identifying brand subgroups

There are several ways of visualizing the results of hierarchical clustering. The best way is to categorize the different objects, in our case, the brands, into a dendogram, which is a type of tree diagram. Fig. 4 shows the dendogram obtained from the clustering analysis. The levels in the dendogram indicate the order in which the clusters emerge. The higher the level of the link between brands, the greater the difference between them. In the 36-brand cluster, for example, "Chosungah 22" and "Giorgio Armani Beauty" are very similar, since the link between them appears at a very low level, but they share less similarity with "NUDESTIX".

In this research, the clustering results are also displayed on the perceptual map resulting from the PCA and shown in Fig. 5. As in the case of the dendogram, it can be seen that cluster 1 contains 36 brands in the blusher category, which might indicate that brands within this category do not differ in terms of the LIWC associations used in this research. Cluster 2 is composed of three brands: Black Up, Estée Lauder and TOM FORD. These brands are perceived by consumers as being more differentiated, although they are relatively close to Cluster 1. Note that Clusters 3 and 4, which are composed of one brand each (Supergoop! and Ciaté London, respectively) are highly differentiated in the

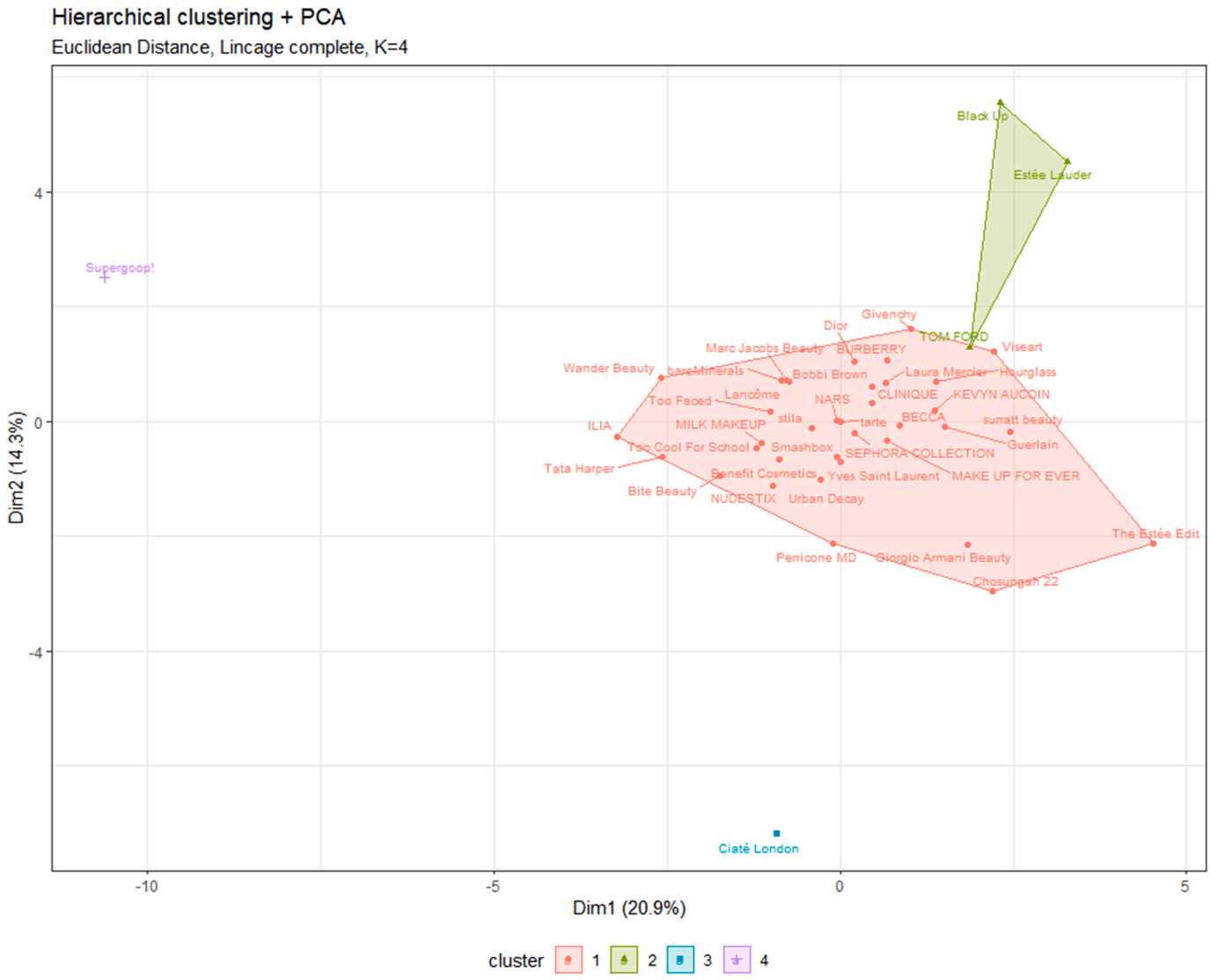


Fig. 5. Brand positioning map and hierarchical clustering results (Agglomerative algorithm: Linkage (Ward's method), k = 4).

market based on these types of psychological associations. For illustrative purposes, Appendix D includes the output when k = 6, from which it can be seen that the blusher market is more segmented, and that a larger number of brand groups could be identified.

Overall, we observe that, in the blusher category, consumers' brand associations are very similar. One possible explanation for this is that blushers vary little in their features from one brand to another. To corroborate this notion, we checked with the online retailer and observed that the product characteristics of every brand in the blusher category are similar in terms of available colours, product packaging, ingredients, etc., which would explain why consumers might perceive them as being so similar to one another. Nevertheless, we should bear in mind that, as claimed by Hair et al. (2010), the cluster solution is not generalizable because it is totally dependent on the variables used as the basis for the similarity measure. Thus, if other LIWC variables are explored, the results might vary. Similarly, the results are dependent upon the type of data used for the clustering analysis. Therefore, online reviews from other product categories will yield different clustering results.

4.3.1. *Describing brand subgroups*

Once the clustering analysis is complete, it is important to describe the characteristics of the clusters in order to detect patterns and identify similarities and differences between them. From the practitioner's viewpoint, an understanding of the differences between brand segments is important when attempting to adjust marketing strategies to consumers' needs and behaviours, based on their perceptions or associations.

Figs. 6 and 7 provide a graphic representation of the composition of each cluster in terms of textual associations and non-textual brand features, respectively. Non-textual brand variables are not used as clustering inputs in our study, because we want to identify segments based on textual brand associations. They are used to describe the clusters, however, because they might be relevant for identifying the specific characteristics of each cluster in terms of non-textual dimensions.

Cluster 1, which includes most of the brands in the category, stands out as having the highest number of online reviews, the highest number of Instagram Followers and the highest sales in the category. In terms of associations, the Cluster 1 brands have average scores on all textual brand associations, which might indicate little differentiation in the market as far as the explored brand perceptions are concerned.

Cluster 2, which is composed of three brands, stands out as the highest priced on average. Cluster 2 has no extreme average scores on the textual brand variables but has by far the closest associations with *Feel*, *Hear* and *See*.

Cluster 3, which includes the brand Ciaté London, has average scores on the non-textual brand dimensions. It has by far the closest associations with *Friend*, *Achievement*, *Time*, *Risk* and *Leisure*.

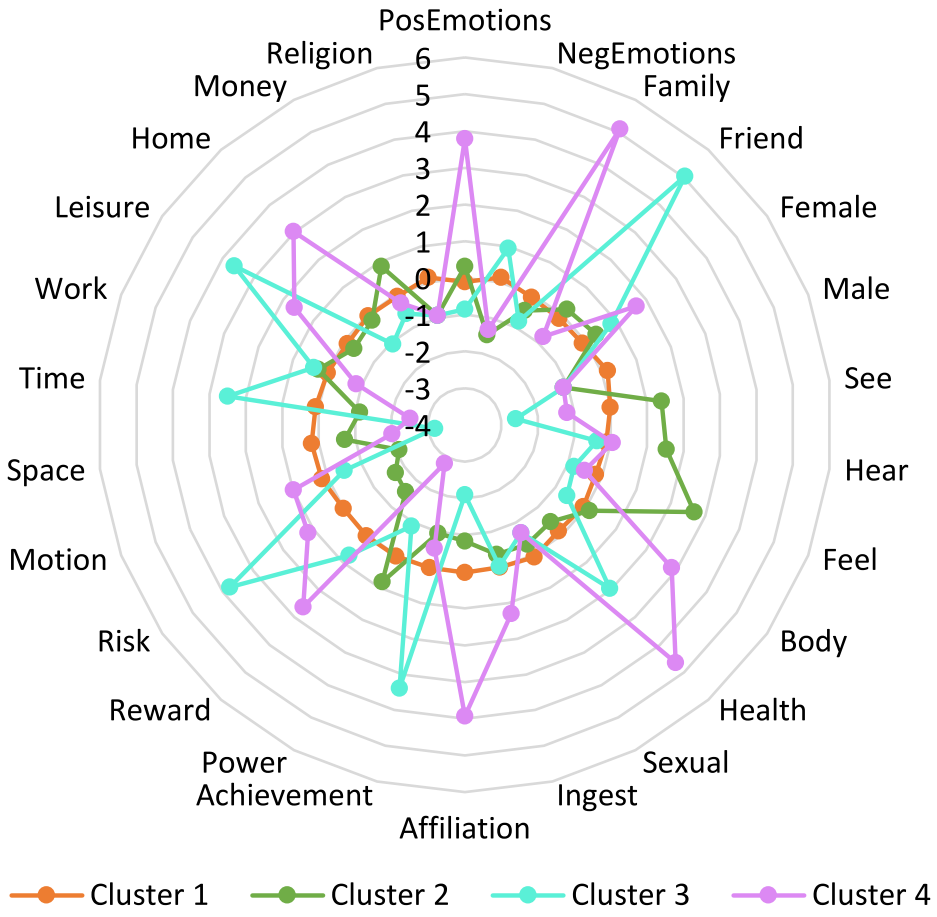


Fig. 6. Cluster descriptions in terms of LIWC text variables.

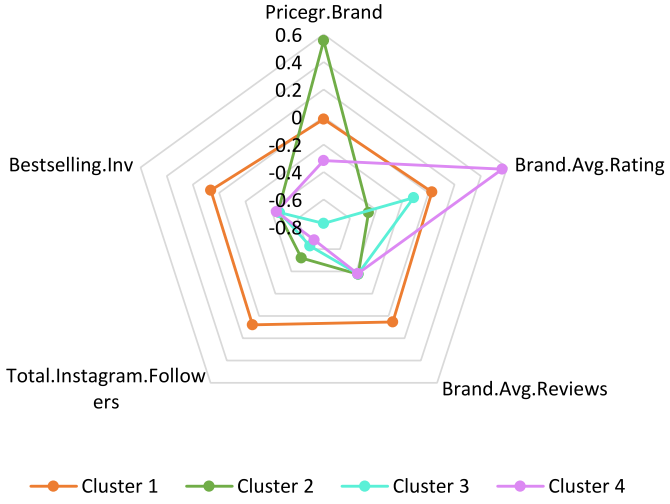


Fig. 7. Cluster descriptions in terms of non-textual brand variables.

Finally, cluster 4 contains only the Supergoop! brand and has the highest average brand rating. In terms of textual brand associations, this cluster is by far the most closely associated with *Family*, *Body*, *Health*, *Affiliation*, *Reward* and *Home*.

5. Conclusions

As a general conclusion, we can say that LIWC is a powerful tool for extracting brand associations from the text of online reviews thanks to the wide range of textual variables it provides. This means that the type of variable adopted in the text mining stage can be selected in accordance with the research objectives. Then, the brand positioning and competition analyses can be conducted with R software packages. We have also observed that the output of the brand positioning analysis is quite product category-dependent. We have noticed that blusher category products are perceived as being quite similar to one another. However, other product categories might show more diversity and differentiation in terms of consumer perceptions.

5.1. Contribution

Although we contextualize this research within the brand image and brand positioning literature, this paper is primarily of a practical nature. Its overall aim is to suggest and illustrate a structured research procedure for text mining aimed at exploring brand image and brand positioning.

We suggest the use of a lexicon-based approach, the LIWC, for text mining, since it eases the research procedure for small or medium sized companies. As well as being affordable and intuitive, the LIWC is also powerful by providing a large set of text variables with which to analyze various aspects of online reviews (or other forms of eWOM) for their relationships with consumer behavior. The literature in this context of brand image and brand positioning has claimed that, when measuring brand associations, consideration should be given not only to physical attributes, but also to functional, emotional and self-expressive benefits (Delgado--Ballester and Fernández-Sabiote, 2016; Ruane and Wallace, 2015; Simon et al., 2016). Accordingly, therefore, we propose the analysis of the set of variables in the LIWC "Psychological processes" group.

There are two main advantages to using online reviews rather than

traditional primary data, that is, mainly survey data and in-depth interviews, for analyzing consumer perceptions. First of all, online reviews are spontaneous (Marchand et al., 2017; Yang and Cho, 2015) because they occur without direct prompting and are usually driven by the consumer's desire to help others or to communicate their status, and are therefore more likely to reflect their true product and brand perceptions. Thus, by analyzing the textual content of online reviews, we are able to identify more genuine perceptions. Furthermore, huge amounts of online reviews can be collected relatively easily. The availability of such large amounts of data allows researchers access to the perceptions of many different consumers, as a result of which their conclusions are clearly stronger than those obtained through qualitative techniques (Reynolds and Gutman, 1979; Rossolatos, 2019; Teichert et al., 2017) including surveys (Baksi and Panda, 2018; Cho et al., 2015; Davis et al., 2009; John et al., 2006; Konuk, 2018).

Moreover, the existing literature on online review text mining is quite diverse in terms of research objectives and choice of text mining techniques. Motivated by this, we propose, explain and illustrate a structured and easy-to-follow procedure for exploring brand image and brand positioning through the text of online reviews.

5.2. Managerial implications

Two types of text mining methods appear in literature, those based on machine learning algorithms and those using a lexicon-based approach. The choice between the two necessarily depends not only on the specific aim of the research but also on the available resources (e.g. skilled personnel, technical infrastructure and data). Magoulas and Swoyer (2020) found that one of the main barriers to Artificial Intelligence (AI) adoption by businesses is that they lack the necessary skills or have difficulty hiring people to fill the required roles. Machine learning is a specific application of AI; therefore text mining analysis based on machine learning algorithms proves too much of a challenge for many companies.

In this scenario, we propose to follow a text mining procedure that relies on LIWC, a lexicon-based text mining method which is easier to implement than machine learning, especially for small and medium companies which might be lacking in available resources and trained personnel. LIWC would allow companies to select the variables of interest to explain how consumers perceive their brands and other brands in the marketplace and identify their main competitors within the market based on different consumer perceptions. Using R software packages, companies can use LIWC output in their product and brand positioning analysis. To enhance the managerial implication of this paper, Appendix A shows the detailed descriptions of steps followed and options selected in R for non-technical users.

Overall, the proposed procedure could enable businesses to obtain insights into aspects such as:

- Detecting whether consumers associate the brand with positive and/or negative emotions, which might lead to a better understanding of customer satisfaction with the brand.
- Understanding how consumers perceive the brand. In this way, companies could increase customer loyalty by reinforcing brand perceptions by means of adapted marketing mix strategies.
- New product development. Insights into the benefits being sought by consumers so as to strengthen some and drop others in future product releases.
- Brand positioning, competition and differentiation. Companies can analyze their brand's positioning in the market, by identifying the main associations. This will show them whether they are positioned as they want to be, and, if not, they can develop strategies to shift their product's positioning to a less saturated area of the perceptual map and thereby differentiate from competitors.

5.3. Limitations and future research

To illustrate the process, we use online reviews for a category of cosmetic products, namely, blushers. Findings from our empirical study are quite context-dependent, since brand associations vary widely from one type of product to another. Therefore, although our proposed research procedure can be used for any type of product or service category, the findings obtained must be evaluated without losing sight of the type of product or service we are dealing with. Moreover, this is an analysis of the broader picture of brand positioning in the blusher category; it is not focused on any specific brand. Individual brands would therefore need to go further by examining their particular cases and thus draw more specific conclusions.

The proposed procedure could be used to analyze the text of online reviews for any type of product, service or brand. It might be interesting to analyze categories of products that drive more varied attribute or quality perceptions than is the case of the blusher category. In the cosmetics industry, for example, the perfume category is likely to include more differentiated products, since fragrances vary substantially between one product and another, and also between brands.

Although this is a brand-level study motivated by our interest in analyzing brand image and brand positioning, the process could be adapted for the analysis of product image and positioning by aggregating online reviews to product-level. It could also be used to identify different consumer segments based on language styles, in which case the aggregation would be at the reviewer-level.

Notwithstanding the widespread use of LIWC in previous text mining literature, future research could use one of the available lexicon-based methods. Machine learning methods for text mining could be also used by companies wishing to extract more specific aspects from online reviews texts, provided they have the necessary resources. In this research, we were especially interested in exploring other brand associations apart from the positive or negative sentiment expressed in online reviews, which is the usual focus in the literature. However, other hidden aspects of texts, perhaps relating to the reviewer's writing style (e.g., informal writing, cognitive writing and time focus), which are also covered by LIWC, could be explored. Another possibility would be to differentiate brand associations in reviews based on review ratings (number of stars). This would reveal whether associations are positive or negative.

In terms of research data, while this research is based on online consumer reviews, the same research procedure could be used with any type of eWOM, such as social networks and blogs. The brand association data for our analysis were obtained through a specific online retailer, so it would be interesting to compare brand associations across different online retailers, to detect possible differences and verify whether they are platform dependent. A further option would be to conduct a survey based on a questionnaire designed to determine whether online brand associations are the same as brand associations in general.

As a final reminder, it is important to bear in mind that the writers of online reviews might not share the same profile as the rest of the consumer population, which might include segments whose opinions are not expressed online. Such consumers might still need to be approached with traditional techniques, such as surveys or in-depth interviews. Ideally, we could combine techniques in order to compare the results in terms of brand image using both online and traditional consumer expression.

Funding sources

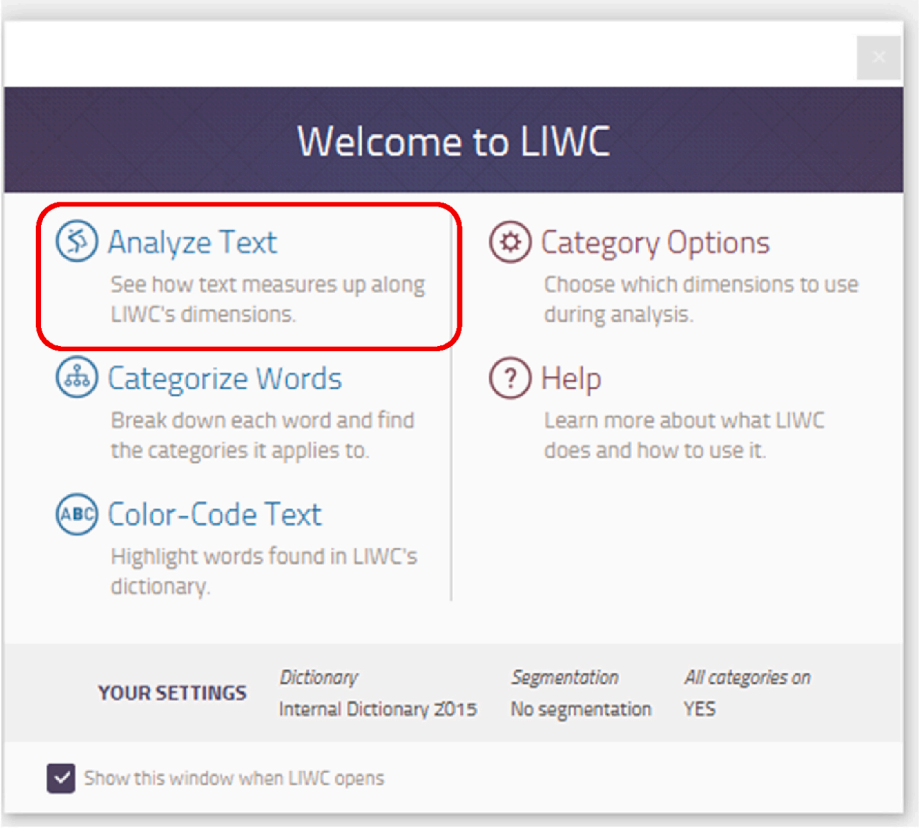
This work was supported by the Spanish Ministry of Economy, Industry and Competitivity [grant number: ECO2015-65393-R] and by the Government of Spain Ministry of Science, Innovation and Universities Grant numbers: ID2019-108554RB-I00.

APPENDIX A. Brand positioning from online reviews: steps and code

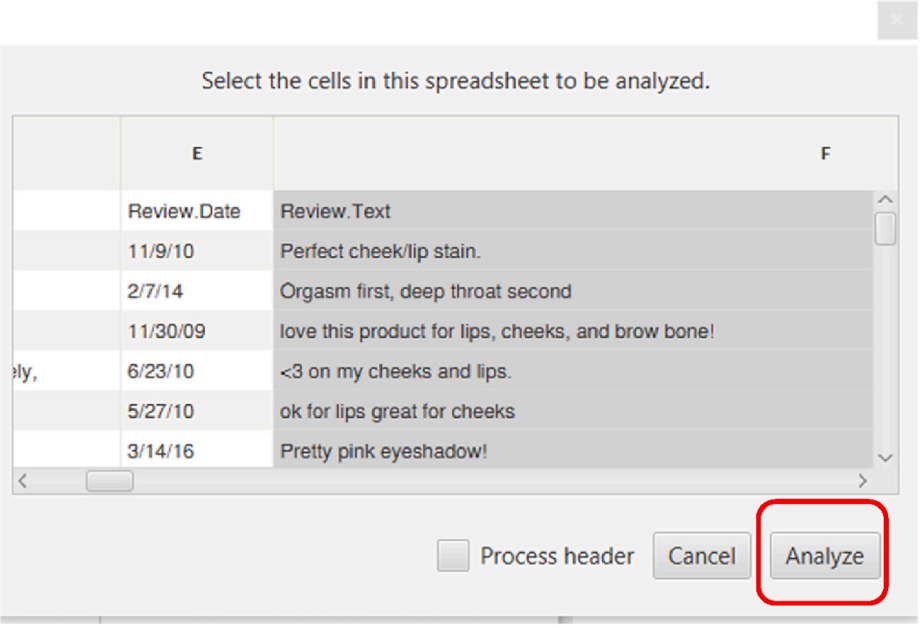
Text Mining

Starting at the LIWC software

1. Importing the CSV file obtained from Web Scraping to the LIWC software.



2. Selecting the column in the csv file containing the text to analyze.



3. Save the output once the analysis is processed into a CSV file.

Data aggregation

Moving to R Studio

From this stage onwards, the analyses are conducted using R Studio, which can be downloaded from https://www.rstudio.com/products/rstudio/. Before running R Studio, the software R should be downloaded and installed from https://cran.r-project.org/mirrors.html.

1. Opening the R packages need for the analyses
   Before starting to run the different codes, it is necessary to install and open the required packages in R.

```
Install.packages(dplyr)
Install.packages (tidyr)
Install.packages (tidytext)
Install.packages (stringi)
Install.packages (stringr)
Install.packages (ggplot2)
Install.packages (scales)
libr Install.packages ary(readr)
Install.packages (tidyverse)
Install.packages (FactoMineR)
Install.packages (factoextra)

library(dplyr)
library(tidyr)
library(tidytext)
library(stringi)
library(stringr)
library(ggplot2)
library(scales)
library(readr)
library(tidyverse)
library(FactoMineR)
library(factoextra)
```

2. Importing the CSV file containing online reviews to R Studio as a data frame

We give the data frame the name of "*Dataframe.Reviews*".

```
Dataframe.Reviews<- read.csv("Path where your CSV file is located on
     your computer\\File Name.csv")
```

3. Selection of reviews having more than 50 words

In *Dataframe. Reviews,* the variable capturing the number of words of online reviews is called "*Review.Length*". Since we decided to keep online reviews having 50 words of more, we should build another data frame, which we name as "*Reviews.50words*", containing those reviews where *Review. Length*>50.

```
Reviews.50words<- subset (Dataframe.Reviews, Dataframe.Reviews$Length
                >50)
```

4. Building a data frame with the variables to be used in the analysis

The CSV file imported into R has many variables obtained from Web scraping, together with the variables provided by LIWC. However, we build a new data frame, with the name of "*Reviews.Analysis*", keeping only those variables of interest for our analyses.

In the "*Reviews.Analysis*" data frame we are keeping the 26 textual variables of interest obtained from the LIWC, the variable "*Product.Brand*", which allows us to know the name of the brand, and four non-textual variables (product price, product bestselling ranking, product average rating, product number of reviews). These non-textual variables allow us to describe our sample and the clusters.

To keep the variables of interest, we should select the corresponding columns to keep from the data frame "Dataframe.Reviews". If our variables of interest are in column 1 to 31, we run:

```
Reviews.Analysis<- Dataframe.Reviews<- [c(1:31)]
```

5. Aggregating variables to the brand-level

So far, the data is recorded at a review-level in the data frame "*Reviews.Analysis*", where every row in the data frame records information about an individual online review. Because we are doing a brand positioning analysis, we are going to work with brand average values, so we should aggregate the variables of interest to the brand-level. Therefore, we should aggregate every variable by the variable "*Product.Brand*".

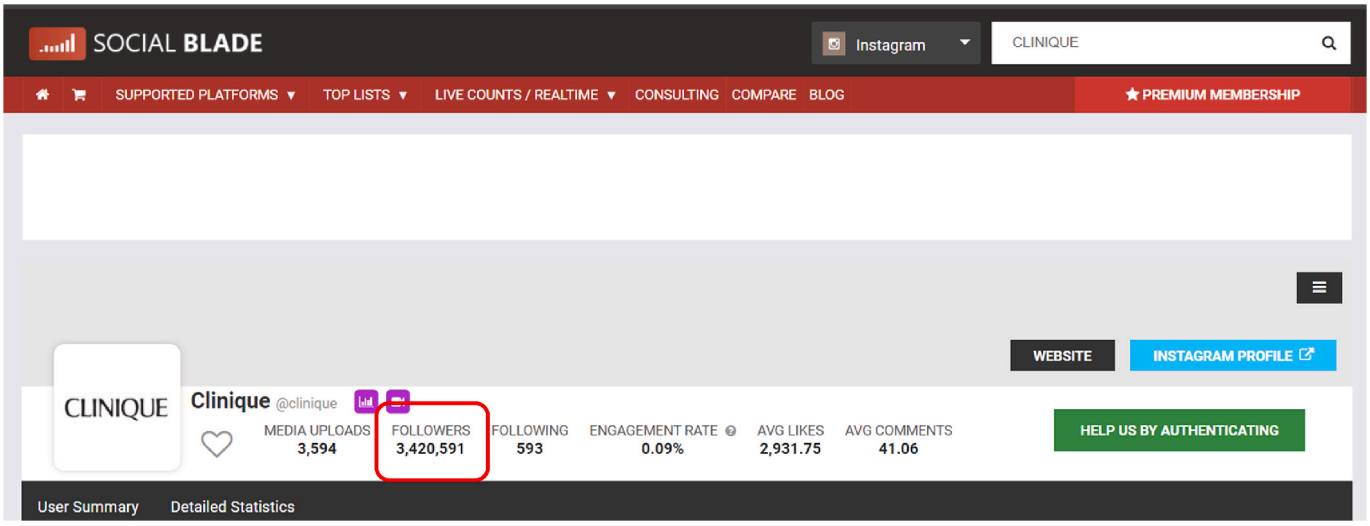
We are going to build a new data frame, with the name "*Brands.Database*", containing brand average values for every variable.

```
Brands.Database<- aggregate(Reviews.Analysis,by=list(Reviews.Analysis
                  $Product.Brand), FUN=mean]
```

6. Merging the data frame with external information (obtained from other sources)

In this research, we collected information about the number of followers of the brands at Instagram, to capture the popularity of the brand at social networks. This information was taken from https://socialblade.com/. To do that, we searched for the number of Instagram followers of each brand in our database on February the 17th 2017. With that information, we built a CSV file with two variables, the brand name "*Product.Brand*" and the number of followers "*Followers*".

Next Figure captures a screenshot of the information provided by Social Blade regarding one of the brands in our database, CLINIQUE. We recorded the information about "*Followers*".



The CSV file containing the information about the number of followers of every brand was imported into R Studio as a data frame.

```
Instagram.Followers<- read.csv("Path where your CSV file is located o
    n your computer\\File Name.csv")
```

Once imported, the data frame "*Instagram.Followers*" is merged with the data frame containing the brand average information "*Brands.Database*". The two data frames are merged by the variable "*Product.Brand*".

```
Brands.Database<-merge(Brands.Database, Instagram.Followers, by="Prod
    uct.Brand")
```

7. Exploring descriptive statistics of variables of interest

Non-textual and textual variables (brand averages) descriptive statistics are explored.

```
Summary(Brands.Database)
```

8. Keeping in the analysis brands having more than 9 reviews

In this research, a cut-off value of 9 reviews was established. Thus, we had to delete from the analysis 3 brands (rms beauty, Shiseido and trèStiQue). To do that, we deleted those rows from the "*Brands.Database*" data frame. In our case those brands were in rows 28,29 and 40.

```
Brands.Database[-c(28,29,40),]
```

Brand positioning: building a perceptual map

9. Using the name of the brand as the row name in the data frame

The PCA should be conducted using only numeric variables. Therefore, before running the PCA code, we should give rows the name of the corresponding brands, which are in the variable "*Product.Brand*".

```
Row.names(Brands.Database)<- Brands.Database$Product.Brand
```

10. Selection of the variables to use in the PCA (LIWC variables)

At this stage, we should select the columns (variables) to keep in our PCA analysis. Let's imagine that the 26 selected LIWC variables which are recorded in "*Dataframe.Reviews*" are in columns 2 to 27, because column 1 belongs to the "*Product.Brand*" variable. We build a new data frame containing only the variables to be used in the PCA.

```
Brands.PCA<-Brands.Database[c(2:27)]
```

The data frame "*Brands.PCA*" is composed by 41 brands and 26 variables (those obtained from the LIWC).

11. Exploring correlations

Before conducting the PCA, we should explore correlations. PCA is conducted when there are correlated variables in the data frame.

```
cor(Brands.PCA[, 1:26])
```

12. Standardizing variables before PCA

Before conducting the PCA, variables should be standardized. We create two databases, one containing only LIWC variables (to be used in the PCA and hierarchical clustering analysis, "*St.Brands.LIWC*") and another data frame containing also non-textual brand information (to be used to describe clusters, "*St.Brands*").

```
St.Brands.LIWC<-scale(Brands.PCA)
St.Brands<-scale(Brands.Database)
```

13. Running the PCA algorithm

Once the variables are standardized, the PCA is conducted using as inputs the 26 textual variables obtained from the LIWC.

```
Brands.PCA<- prcomp(St.Brands.LIWC[,c(1:26)], center = TRUE,scale. = T
RUE)
```

13.1 Looking at the explained variance of principal components (PCs)

We have to explore the variance explained by PCs.

```
Summary(Brands.PCA)
```

13.2 Exploring PCs loadings

Now, we should explore the loadings of the PC on the different textual variables.

```
Brands.PCA$rotation
```

13.3 Exploring brands' loadings

Loadings are also calculated for each brand on each principal component, which represents the input to draw the perceptual map.

```
Head(Brands.PCA$x)
```

13.4 Plotting the perceptual map on PC1 & PC2

There are several plotting options. First, we can plot only the brands.

```
fviz_pca_ind(Brands.PCA, repel=TRUE)
```

Second, we can represent a combination of brands and textual variables, showing with arrows the direction and magnitude of each textual variable, and with points the positioning of each brand.

```
var<-get_pca_var(Brands.PCA)

fviz_pca_biplot(Brands.PCA ,repel = TRUE,col.var = "#2E9FDF",col.ind =
"#696969")
```

13.5 Plotting perceptual maps on other PCs combinations

In this research, we also present the perceptual map on PC1&PC3 and on PC2&PC3.

```
fviz_pca_ind(Brands.PCA,repel = TRUE,col.var = "#2E9FDF",col.ind = "#6
96969",axes = c(1,3))

fviz_pca_ind(Brands.PCA,repel = TRUE,col.var = "#2E9FDF",col.ind = "#6
96969",axes = c(2,3))
```

14. Brand positioning: identifying brand subgroups

14.1 Establishing the optimal number of clusters

We used the Elbow method to establish the optimal number of clusters. The data frame used for the analysis is the one having the textual variables of interest standardized, "*St.Brands*".

```
fviz_nbclust(St.Brands.LIWC, kmeans, method = "wss") +
  geom_vline(xintercept = 4, linetype = 2)+
  labs(subtitle = "Elbow method")
```

14.2 Plotting the dendogram

In this research, we adopted a hierarchical clustering technique, in particular an agglomerative method using the Wald's method clustering algorithm.

```
dist <- dist(St.Brands.LIWC)
seg.hc <- hclust(dist, method="ward.D2")
fviz_dend(x = seg.hc, k = 6, cex = 0.6) +
  geom_hline(yintercept = 5.5, linetype = "dashed") +
  labs(title = "Herarchical clustering",
       subtitle = "Euclidean distance, Lincage Ward, K=4")
```

14.3 Plotting the clustering output into the perceptual map

```
fviz_cluster(object = list(data=St.Brands.LIWC, cluster=cutree(seg.hc, k=4)),
             ellipse.type = "convex", repel = TRUE, show.clust.cent = FALSE,
             labelsize = 8)  +
  labs(title = "Hierarchical clustering + PCA",
       subtitle = "Euclidean Distance, Linkage complete, K=4") +
  theme_bw() +
  theme(legend.position = "bottom")
```

14.4 Describing clusters' characteristics

We describe clusters based on non-textual characteristics (brand average price, brand average rating, brand bestselling ranking, brand average number of reviews per product and total Instagram followers) and on textual characteristics. For that, we use the data frame containing non-textual and textual standardized variables at a brand-level, "*St.Brands*".

```
as.data.frame(St.Brands) %>% mutate(Cluster = final$cluster) %>% group
_by(Cluster) %>% summarise_all("mean") %>% kable() %>% kable_styling()
```

APPENDIX B. Complete List of Psychological Processes Variables provided by the LIWC

Affective Processes
Positive emotion
Negative emotion
Anxiety
Anger
Sadness
Social processes
Family
Friends
Female references
Male references
Cognitive processes
Insight
Causation
Discrepancy
Tentative
Certainty
Differentiation
Perceptual processes
See
Hear
Feel
Biological processes
Body
Health
Sexual
Ingestion
Drives
Affiliation
Achievement
Power
Reward
Risk
Time orientation
Past focus
Present focus
Future focus
Relativity
Motion
Space
Time
Personal concerns
Work
Leisure
Home
Religion
Death
Informal language
Swear words
Netspeak
Assent
Nonfluencies
Fillers

APPENDIX C. PCA Maps on PC1&PC3 and on PC2&PC3

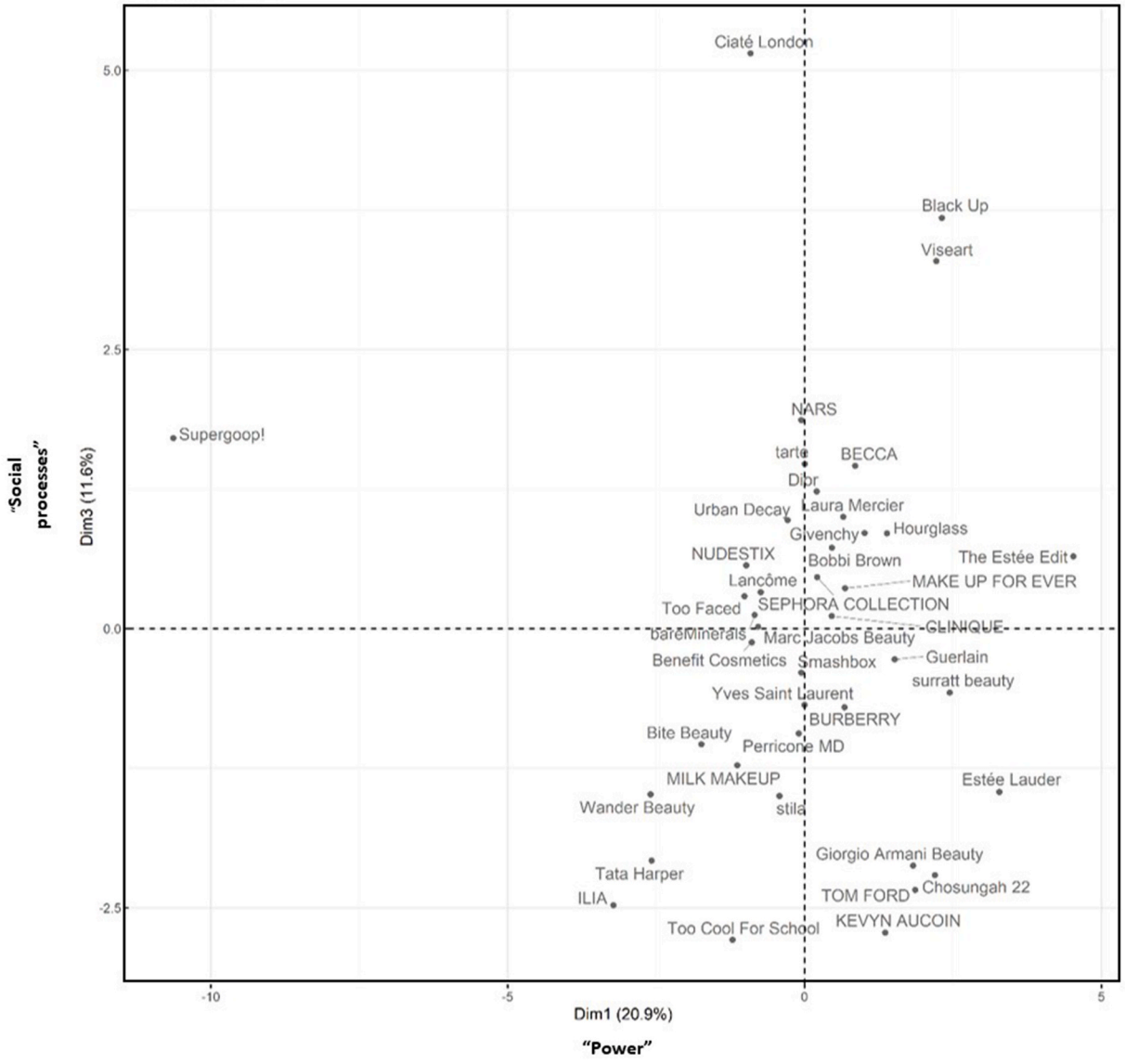


Fig. A1. Two-dimensional brand positioning map on PC1 & PC3.

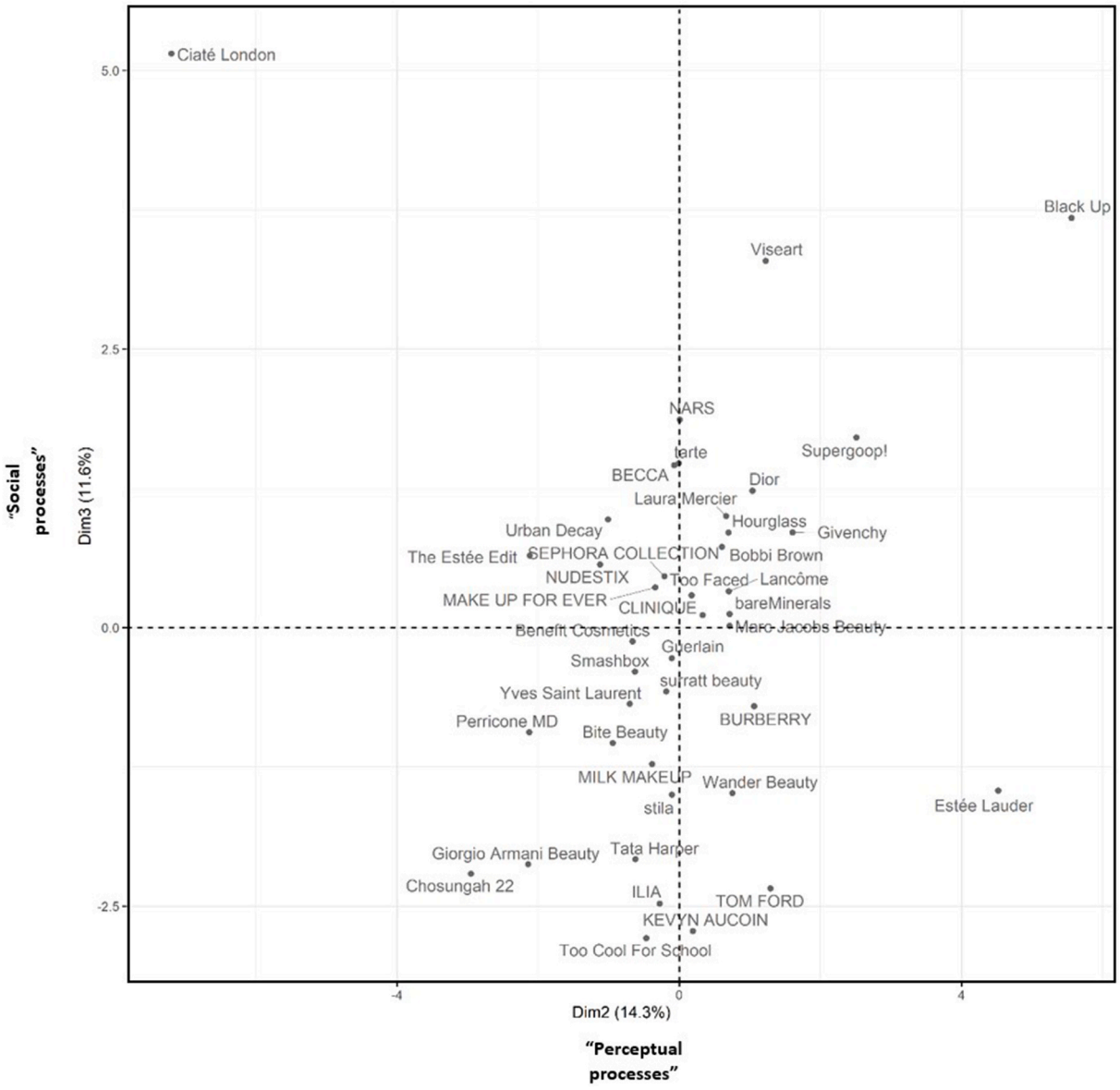
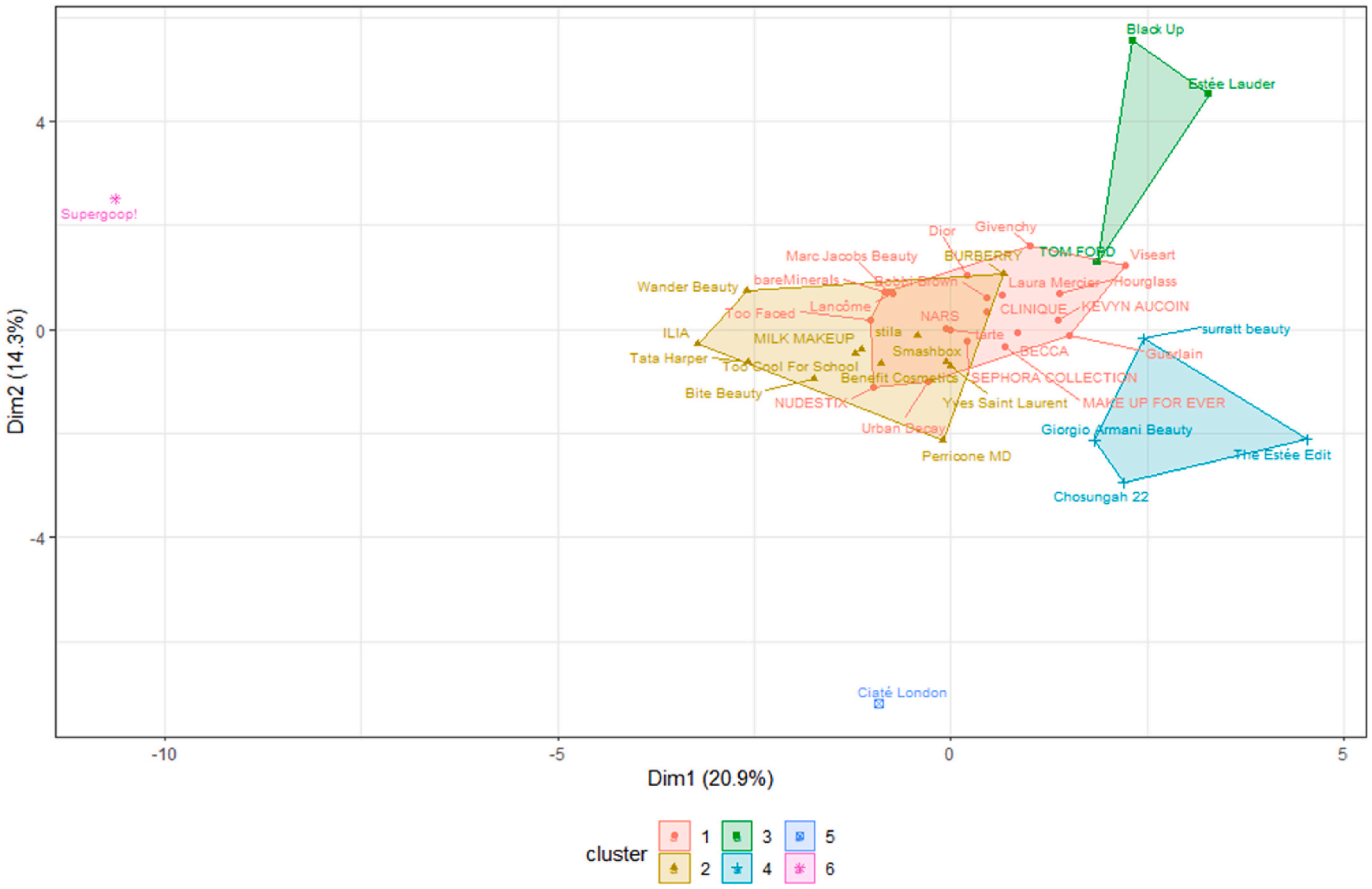


Fig. A2. Two-dimensional brand positioning map on PC2&PC3.

APPENDIX D. Hierarchical Clustering (K = 6)



References

Aaker, D.A., 1991. Managing Brand Equity. Manag. Brand Equity.

Ahani, A., Nilashi, M., Yadegaridehkordi, E., Sanzogni, L., Tarik, A.R., Knox, K., Samad, S., Ibrahim, O., 2019. Revealing customers' satisfaction and preferences through online review analysis: the case of Canary Islands hotels. J. Retailing Consum. Serv. 51, 331–343. https://doi.org/10.1016/j.jretconser.2019.06.014.

Anselmsson, J., Burt, S., Tunca, B., 2017. An integrated retailer image and brand equity framework: Re-examining, extending, and restructuring retailer brand equity. J. Retailing Consum. Serv. 38, 194–203. https://doi.org/10.1016/j.jretconser.2017.06.007.

Areni, C.S., 2003. The effects of structural and grammatical variables on persuasion: an elaboration likelihood model perspective. Psychol. Market. 20, 349–375. https://doi.org/10.1002/mar.10077.

Baksi, A.K., Panda, T.K., 2018. Branding destinations with multisensory brand associations and evaluating its impact on behavioural pattern under the intervention of multiplex phenomenon of relationship-branding. Manag. Sci. Lett. 8, 1169–1182. https://doi.org/10.5267/j.msl.2018.8.007.

Balducci, B., Marinova, D., 2018. Unstructured data in marketing. J. Acad. Market. Sci. https://doi.org/10.1007/s11747-018-0581-x.

Berger, J., Humphreys, A., Ludwig, S., Moe, W.W., Netzer, O., Schweidel, D.A., 2020. Uniting the tribes: using text for marketing insight. J. Market. 84, 1–25. https://doi.org/10.1177/0022242919873106.

Bholowalia, P., Kumar, A., 2014. EBK-means: a clustering technique based on Elbow method and K-means in WSN. Int. J. Comput. Appl. 105, 975–8887.

Bijmolt, T.H.A., Wedel, M., DeSarbo, W.S., 2021. Adaptive multidimensional scaling: brand positioning based on decision sets and dissimilarity judgments. Cust. Needs Solut. 8, 1–15. https://doi.org/10.1007/s40547-020-00112-7.

Boyd, R.L., 2017. Data analytics in digital humanities. Data Anal. Digit. Humanit. https://doi.org/10.1007/978-3-319-54499-1.

Brandt, C., De Mortanges, C.P., Bluemelhuber, C., Van Riel, A.C.R., 2011. Associative networks: a new approach to market segmentation. Int. J. Mark. Res. 53, 187–207. https://doi.org/10.2501/IJMR-53-2-187-208.

Büyükdağ, N., Kitapci, O., 2021. Antecedents of consumer-brand identification in terms of belonging brands. J. Retailing Consum. Serv. 59 https://doi.org/10.1016/j.jretconser.2020.102420.

Chang, R.C.Y., Mak, A.H.N., 2018. Understanding gastronomic image from tourists' perspective: a repertory grid approach. Tourism Manag. 68, 89–100. https://doi.org/10.1016/j.tourman.2018.03.004.

Chen, K., Kou, G., Shang, J., Chen, Y., 2015. Visualizing market structure through online product reviews: integrate topic modeling, TOPSIS, and multi-dimensional scaling approaches. Electron. Commer. Res. Appl. 14, 58–74. https://doi.org/10.1016/j.elerap.2014.11.004.

Cheng-Hsui Chen, A., 2001. Using free association to examine the relationship between the characteristics of brand associations and brand equity. J. Prod. Brand Manag. 10, 439–451. https://doi.org/10.1108/10610420110410559.

Chevalier, J., Mayzlin, D., 2006. The effect of word of mouth on sales: online book reviews. J. Market. Res. 43, 345–354. https://doi.org/10.1509/jmkr.43.3.345.

Cho, E., Fiore, A.M., Russell, W., D, 2015. Validation of a fashion brand image scale capturing cognitive, sensory, and affective associations: testing its role in an extended brand equity model. Psychol. Market. 32, 28–48. https://doi.org/10.1002/mar.

Chung, C.K., Pennebaker, J.W., 2007. The psychological function of function words. Soc. Commun. Front. Soc. Psychol. 343–359.

Cohn, M.a., Mehl, M.R., Pennebaker, J.W., 2001. Markers of linguistic psychological change surrounding september 11, 2001. Psychol. Sci. 15, 687–693. https://doi.org/10.1111/j.0956-7976.2004.00741.x.

Culotta, A., Cutler, J., 2016. Mining brand perceptions from twitter social networks. Market. Sci. 35, 343–362. https://doi.org/10.1287/mksc.2015.0968.

Cunningham, J.P., Ghahramani, Z., 2015. Linear dimensionality reduction: survey, insights, and generalizations. J. Mach. Learn. Res. 16, 2859–2900.

Davis, D.F., Golicic, S.L., Marquardt, A., 2009. Measuring brand equity for logistics services. Int. J. Logist. Manag. 20, 201–212. https://doi.org/10.1108/09574090910981297.

Delgado-Ballester, E., Fernández-Sabiote, E., 2016. Once upon a brand": storytelling practices by Spanish brands. Spanish J. Mark. - ESIC 20, 115–131. https://doi.org/10.1016/j.sjme.2016.06.001.

DeSarbo, W.S., Park, J., Rao, V.R., 2011. Deriving joint space positioning maps from consumer preference ratings. Market. Lett. 22 (1), 1–14. https://doi.org/10.1007/s11002-009-9100-7.
Filieri, R., Hofacker, C.F., Alguezaui, S., 2018. What makes information in online consumer reviews diagnostic over time? The role of review relevancy, factuality, currency, source credibility and ranking score. Comput. Hum. Behav. 80, 122–131. https://doi.org/10.1016/j.chb.2017.10.039.
France, S.L., Ghose, S., 2016. An analysis and visualization methodology for identifying and testing market structure. Market. Sci. 35, 182–197. https://doi.org/10.1287/mksc.2015.0958.
Gensler, S., Völckner, F., Egger, M., Fischbach, K., Schoder, D., 2015. Listen to your customers: insights into brand image using online consumer-generated product reviews. Int. J. Electron. Commer. 20, 112–141. https://doi.org/10.1080/10864415.2016.1061792.
Girard, T., Dion, P., 2010. Validating the search, experience, and credence product classification framework. J. Bus. Res. 63, 1079–1087. https://doi.org/10.1016/j.jbusres.2008.12.011.
Guo, Y., Barnes, S.J., Jia, Q., 2017. Mining meaning from online ratings and reviews: tourist satisfaction analysis using latent dirichlet allocation. Tourism Manag. 59, 467–483. https://doi.org/10.1016/j.tourman.2016.09.009.
Gwin, C.F., Gwin, C.R., 2003. Product attributes model: a tool for evaluating brand positioning. J. Market. Theor. Pract. 11, 30–42. https://doi.org/10.1080/10696679.2003.11658494.
Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., 2010. Multivariate Analysis, seventh ed. 7 th Edition.
Hamby, A., Daniloski, K., Brinberg, D., 2015. How consumer reviews persuade through narratives. J. Bus. Res. 68, 1242–1250. https://doi.org/10.1016/j.jbusres.2014.11.004.
Hartmann, J., Huppertz, J., Schamp, C., Heitmann, M., 2019. Comparing automated text classification methods. Int. J. Res. Market. 36, 20–38. https://doi.org/10.1016/j.ijresmar.2018.09.009.
Hauser, J.R., Koppelman, F.S., 1979. Alternative perceptual mapping techniques : relative accuracy and usefulness. J. Market. Res. 16, 495–506.
Henderson, G.R., Iacobucci, D., Calder, B.J., 1998. Brand diagnostics: mapping branding effects using consumer associative networks. Eur. J. Oper. Res. 111, 306–327. https://doi.org/10.1016/S0377-2217(98)00151-9.
Heng, Y., Gao, Z., Jiang, Y., Chen, X., 2018. Exploring hidden factors behind online food shopping from Amazon reviews: a topic mining approach. J. Retailing Consum. Serv. 42, 161–168. https://doi.org/10.1016/j.jretconser.2018.02.006.
Hofmann, J., Clement, M., Völckner, F., Hennig-Thurau, T., 2017. Empirical generalizations on the impact of stars on the economic success of movies. Int. J. Res. Market. 34, 442–461. https://doi.org/10.1016/j.ijresmar.2016.08.006.
Holtgraves, T., Lasky, B., 1999. Linguistic power and persuasion. J. Lang. Soc. Psychol. 18, 196–205. https://doi.org/10.1177/0261927X99018002004.
Hu, F., Trivedi, R.H., 2020. Mapping hotel brand positioning and competitive landscapes by text-mining user-generated content. Int. J. Hospit. Manag. 84, 102317 https://doi.org/10.1016/j.ijhm.2019.102317.
Ireland, M.E., Slatcher, R.B., Eastwick, P.W., Scissors, L.E., Finkel, E.J., Pennebaker, J. W., 2011. language style matching predicts relationship initiation and stability. Psychol. Sci. 22, 39–44. https://doi.org/10.1177/0956797610392928.
Jalilvand, M.R., Samiei, N., 2012. The effect of electronic word of mouth on brand image and purchase intention: an empirical study in the automobile industry in Iran. Market. Intell. Plann. 30, 460–476. https://doi.org/10.1108/02634501211231946.
John, D.R., Loken, B., Kim, K., Monga, A.B., 2006. Brand concept maps: a methodology for identifying brand association networks. J. Market. Res. 43, 549–563. https://doi.org/10.1509/jmkr.43.4.549.
Kassambara, A., Mundt, F., 2020. Factoextra. Extract and Visualize the Results of Multivariate Data Analyses.
Kaufman, L., Rousseeauw, P.J., 2009. Finding Groups in Data: an Introduction to Cluster Analysis.
Kawaf, F., Istanbulluoglu, D., 2019. Online fashion shopping paradox: the role of customer reviews and facebook marketing. J. Retailing Consum. Serv. 48, 144–153. https://doi.org/10.1016/j.jretconser.2019.02.017.
Kelly, G.A., 1991. The Psychology of Personal Constructs.
Kim, H.B., Kim, W.G., An, J.A., 2003. The effect of consumer-based brand equity on firms' financial performance. J. Consum. Market. 20, 335–351. https://doi.org/10.1108/07363760310483694.
Kim, S.G., Kang, J., 2018. Analyzing the discriminative attributes of products using text mining focused on cosmetic reviews. Inf. Process. Manag. 54, 938–957. https://doi.org/10.1016/j.ipm.2018.06.003.
King, M.F., Balasubramanian, S.k., 1994. The effects of expertise, end goal, and product type on adoption of preference formation strategy. J. Acad. Market. Sci. 22.
Konuk, F.A., 2018. The role of store image, perceived quality, trust and perceived value in predicting consumers' purchase intentions towards organic private label food. J. Retailing Consum. Serv. 43, 304–310. https://doi.org/10.1016/j.jretconser.2018.04.011.
Kotler, P., Armstrong, G., 2020. Principles of Marketing, 18th Globa. Pearson Education Limited.
Kozinets, R.V., De Valck, K., Wojnicki, A.C., Wilner, S.J.S., 2010. Networked narratives: understanding word-of-mouth marketing in online communities. J. Market. 74, 71–89. https://doi.org/10.1509/jmkg.74.2.71.
Kübler, R.V., Colicev, A., Pauwels, K.H., 2019. Social media's impact on the consumer mindset: when to use which sentiment extraction tool? J. Interact. Market. 50, 136–155. https://doi.org/10.1016/j.intmar.2019.08.001.
Kudeshia, C., Kumar, A., 2017. Social eWOM: does it affect the brand attitude and purchase intention of brands? Manag. Res. Rev. 40, 310–330. https://doi.org/10.1108/MRR-07-2015-0161.
Lee, T.Y., Bradlow, E.T., 2011. Automated marketing research using online customer reviews. J. Market. Res. 48, 881–894. https://doi.org/10.1509/jmkr.48.5.881.
Li, S.T., Pham, T.T., Chuang, H.C., 2019. Do reviewers' words affect predicting their helpfulness ratings? Locating helpful reviewers by linguistics styles. Inf. Manag. 56, 28–38. https://doi.org/10.1016/j.im.2018.06.002.
Li, X., Hitt, L.M., 2008. Self-selection and information role of online product reviews. Inf. Syst. Res. 19, 456–474. https://doi.org/10.1287/isre.1070.0154.
Liu, X., Burns, A.C., Hou, Y., 2017. An investigation of brand-related user-generated content on twitter. J. Advert. 46, 236–247. https://doi.org/10.1080/00913367.2017.1297273.
Londoño, J.C., Elms, J., Davies, K., 2016. Conceptualising and measuring consumer-based brand-retailer-channel equity. J. Retailing Consum. Serv. 29, 70–81. https://doi.org/10.1016/j.jretconser.2015.11.004.
Low, G.S., Lamb, C.W., 2000. The measurement and dimensionality of brand associations. J. Prod. Brand Manag. 9, 350–370. https://doi.org/10.1108/10610420010356966.
Ludwig, S., de Ruyter, K., Friedman, M., Brüggen, E.C., Wetzels, M., Pfann, G., 2013. More than words: the influence of affective content and linguistic style matches in online reviews on conversion rates. J. Market. 77, 87–103. https://doi.org/10.1509/jm.11.0560.
Magoulas, R., Swoyer, S., 2020. AI adoption in the enterprise 2020. https://doi.org/10.1017/CBO9781107415324.004.
Mahr, D., Stead, S., Odekerken-Schröder, G., 2019. Making sense of customer service experiences: a text mining review. J. Serv. Market. 33, 88–103. https://doi.org/10.1108/JSM-10-2018-0295.
Maimon, O., 2005. Clustering methods. In: Data Mining and Knowledge Discovery Handbook. Springer, Boston, pp. 321–352. https://doi.org/10.1007/978-3-642-93155-0_7.
Malhotra, N.K., 1981. A scale to measure self-concepts, person concepts, and product concepts. J. Market. Res. 18, 456. https://doi.org/10.2307/3151339.
Marchand, A., Hennig-Thurau, T., Wiertz, C., 2017. Not all digital word of mouth is created equal: understanding the respective impact of consumer reviews and microblogs on new product success. Int. J. Res. Market. 34, 336–354. https://doi.org/10.1016/j.ijresmar.2016.09.003.
Martínez Salinas, E., Pina Pérez, J.M., 2009. Modeling the brand extensions' influence on brand image. J. Bus. Res. 62 (1), 50–60. https://doi.org/10.1016/j.jbusres.2008.01.006.
Mohammad, S., Turney, P., 2010. Emotions evoked by common words and phrases. Proc. NAACL-HLT 26–34.
Moon, S., Jalali, N., Erevelles, S., 2021. Segmentation of both reviewers and businesses on social media. J. Retailing Consum. Serv. 61, 102524 https://doi.org/10.1016/j.jretconser.2021.102524.
Moon, S., Kamakura, W.A., 2017. A picture is worth a thousand words: translating product reviews into a product positioning map. Int. J. Res. Market. 34, 265–285. https://doi.org/10.1016/j.ijresmar.2016.05.007.
Nasiri, M.S., Shokouhyar, S., 2021. Actual consumers' response to purchase refurbished smartphones: exploring perceived value from product reviews in online retailing. J. Retailing Consum. Serv. 62, 102652 https://doi.org/10.1016/j.jretconser.2021.102652.
Nelson, P., 1970. Information and Consumer Behavior, 78, pp. 311–329.
Netzer, O., Feldman, R., Goldenberg, J., Fresko, M., 2012. Mine your own business: market-structure surveillance through text mining. Market. Sci. 31, 521–543. https://doi.org/10.1287/mksc.1120.0713.
Olson, J.C., Muderrisoglu, A., 1979. The stability of responses obtained by free elicitation: implications for measuring attribute salience and memory structure. Adv. Consum. Res. 6, 269–275.
Panda, S., Pandey, S.C., Bennett, A., Tian, X., 2019. University brand image as competitive advantage: a two-country study. Int. J. Educ. Manag. 33 (2), 234–251. https://doi.org/10.1108/IJEM-12-2017-0374.
Park, H.-J., Rabolt, J.N., 2009. Cultural value, consumption value, and global brand image: a cross-national study. Psychol. Market. 26, 714–735. https://doi.org/10.1002/mar.
Park, H.E., Yap, S.F.C., Makkar, M., 2019. A laddering study of motivational complexities in mobile shopping. Market. Intell. Plann. 37, 182–196. https://doi.org/10.1108/MIP-03-2018-0104.
Peladeau, N., 2016. WordStat: Content Analysis Module for SIM- STAT.
Pennebaker, J.W., Boyd, R., Jordan, K., Blackburn, K., 2015. The Development and Psychometric Properties of LIWC2015. Austin, TX Univ. Texas Austin 1–22. https://doi.org/10.15781/T29G6Z.
Pennebaker, J.W., Chung, C.K., Ireland, M., Gonzales, A., Booth, R.J., 2007. The Development and Psychometric Properties of LIWC2007. Austin, Tx LIWC.Net 1. https://doi.org/10.1177/026377588300100203.
Pennebaker, J.W., Francis, M.E., 1996. Cognitive, emotional, and language processes in disclosure. Cognit. Emot. 10, 601–626. https://doi.org/10.1080/026999396380079.
Puranam, D., Narayan, V., Kadiyali, V., 2017. The effect of calorie posting regulation on consumer opinion: a flexible latent dirichlet allocation model with informative priors. Market. Sci. 36, 726–746. https://doi.org/10.1287/mksc.2017.1048.
Reynolds, T.J., Gutman, J., 1979. Laddering theory, method, analysis, and interpretation. J. Advert. Res. 28, 11–31.
Rossolatos, G., 2019. Negative brand meaning co-creation in social media brand communities: a laddering approach using NVivo. Psychol. Market. 36, 1249–1266. https://doi.org/10.1002/mar.21273.

Ruane, L., Wallace, E., 2015. Brand tribalism and self-expressive brands: social influences and brand outcomes. J. Prod. Brand Manag. 24, 333–348. https://doi.org/10.1108/JPBM-07-2014-0656.
Samuels, P., 2016. Advice on Exploratory Factor Analysis. Cent. Acad. Success, vol. 2. Birmingham City Univ.
Schnittka, O., Sattler, H., Zenker, S., 2012. Advanced brand concept maps: a new approach for evaluating the favorability of brand association networks. Int. J. Res. Market. 29, 265–274. https://doi.org/10.1016/j.ijresmar.2012.04.002.
Senecal, S., Nantel, J., 2004. The influence of online product recommendations on consumers' online choices. J. Retailing 80, 159–169. https://doi.org/10.1016/j.jretai.2004.04.001.
Sheldon, P., Bryant, K., 2016. Instagram: motives for its use and relationship to narcissism and contextual age. Comput. Hum. Behav. 58, 89–97. https://doi.org/10.1016/j.chb.2015.12.059.
Simon, C., Brexendorf, T.O., Fassnacht, M., 2016. The impact of external social and internal personal forces on consumers' brand community engagement on Facebook. J. Prod. Brand Manag. 25, 409–423. https://doi.org/10.1108/JPBM-03-2015-0843.
Social Blade [WWW Document], 2017. URL https://socialblade.com/.
Spears, N., Singh, S.N., 2004. Measuring attitude toward the brand and purchase intentions. J. Curr. Issues Res. Advert. 26, 53–66. https://doi.org/10.1080/10641734.2004.10505164.
SproutSocial, 2018. Instagram stats [WWW Document]. URL. https://sproutsocial.com/insights/instagram-stats/.
Swoboda, B., Weindel, J., Hälsig, F., 2016. Predictors and effects of retail brand equity - a cross-sectoral analysis. J. Retailing Consum. Serv. 31, 265–276. https://doi.org/10.1016/j.jretconser.2016.04.007.
Tausczik, Y.R., Pennebaker, J.W., 2010. The psychological meaning of words: LIWC and computerized text analysis methods. J. Lang. Soc. Psychol. 29, 24–54. https://doi.org/10.1177/0261927X09351676.
Teichert, T., Gainsbury, S.M., Mühlbach, C., 2017. Positioning of online gambling and gaming products from a consumer perspective: a blurring of perceived boundaries. Comput. Hum. Behav. 75, 757–765. https://doi.org/10.1016/j.chb.2017.06.025.
Tirunillai, S., Tellis, G.J., 2014. Mining marketing meaning from online chatter: strategic brand analysis of big data using latent dirichlet allocation. J. Market. Res. 51, 463–479. https://doi.org/10.1509/jmr.12.0106.
Urban, G.L., Johnson, P.L., Hauser, J.R., 1984. Testing competitive market structures. Market. Sci. 3, 83–112.
Veloutsou, C., Delgado-Ballester, E., 2018. New challenges in brand management. Spanish J. Mark. - ESIC 22, 255–272. https://doi.org/10.1108/SJME-12-2018-036.
Verma, S., Yadav, N., 2021. Past, present, and future of electronic word of mouth (EWOM). J. Interact. Market. 53, 111–128. https://doi.org/10.1016/j.intmar.2020.07.001.
Walker, M.A., Mehl, M.R., Moore, R.K., Mairesse, F., 2007. Using linguistic cues for the automatic recognition of personality in conversation and text. J. Artif. Intell. Res. 30, 457–500.
Wang, W., Feng, Y., Dai, W., 2018. Topic analysis of online reviews for two competitive products using latent Dirichlet allocation. Electron. Commer. Res. Appl. 29, 142–156. https://doi.org/10.1016/j.elerap.2018.04.003.
Webster, G.D., Blanton, H., LaCroix, M., J, M., 2019. Measurement in Social Psychology, Measurement in Social Psychology. Taylor & Francis, Oxon. https://doi.org/10.4324/9780429452925.
Whyte, L.J., 2018. Eliciting cruise destination attributes using repertory grid analysis. J. Destin. Mark. Manag. 10, 172–180. https://doi.org/10.1016/j.jdmm.2018.11.003.
Wong, C.U.I., Qi, S., 2017. Tracking the evolution of a destination's image by text-mining online reviews - the case of Macau. Tourism Manag. Perspect. 23, 19–29. https://doi.org/10.1016/j.tmp.2017.03.009.
Yang, H., Cho, S., 2015. Understanding Brands with Visualization and Keywords from eWOMusing Distributed Representation.
Zhang, J., 2019. What's yours is mine: exploring customer voice on Airbnb using text-mining approaches. J. Consum. Market. 36, 655–665. https://doi.org/10.1108/JCM-02-2018-2581.