ARTICLE IN PRESS

Technological Forecasting & Social Change xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

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

Does country-of-origin brand personality generate retail customer lifetime value? A Big Data analytics approach

Lan-Lung (Luke) Chiang[a,*], Chin-Sheng Yang[b]

[a] College of Management, Yuan Ze University, Chungli, Taoyuan City, Taiwan
[b] Department of Information Management, Yuan Ze University, Chungli, Taoyuan City, Taiwan

ARTICLE INFO

Keywords:
Big Data analytics
Customer-driven marketing strategy
Country-of-origin
Brand personality
Customer lifetime value
Retail industry

ABSTRACT

Many retail firms have witnessed the erosion of customer loyalty with the rise of e-commerce and its resulting benefits to consumers, including increased choices, lower prices, and ease of brand switching. Retailers have long collected data to learn about customer purchasing habits; however, many currently do not use data-mining analytics to increase marketing effectiveness by predicting future buying patterns and potential customer lifetime value, particularly to important segments such as loyal and potential repeat customers. Data mining can efficiently analyze large amounts of business data ("Big Data") in an effort to forecast consumer needs and increase the lifetime value of customers (CLV). Previous studies on these topics primarily focus on conceptual assumptions and generally do not present empirically valid models.

The present study sought to fill the research gap by using Big Data analytics to analyze approximately 44,000 point-of-sale transaction records for 26,000 customers of a Taiwanese retail store to understand how consumer personality traits relate to the country-of-origin (COO) traits (brand personality) of beer brands, and to predict potential customer lifetime value (CLV). The findings revealed that consumers tend to purchase and co-purchase brands with traits similar to their own personality traits (i.e., Japan—peacefulness, Belgium—openness, Ireland—excitement, etc.). Significantly, customers with the group of personality traits associated with "peacefulness" and "openness" were the most profitable customers among the five analyzed clusters (CLV value = 0.3149, 0.2635). The study provides valuable new insights into COO brand personality and consumer personality traits with co-purchase behaviors via data mining techniques, and highlights the value of extending CLV in developing useful marketing strategies.

1. Introduction

Competition is both intensifying and shifting to new arenas in the current retail business environment in response to the increase in comparable and competitive offerings of products and services and growing consumer options (Bharadwaj et al., 2013). It has been repeatedly indicated in the literature that consumers are rapidly evolving their approaches to making purchase decisions, often resulting in the erosion of consumer loyalty (Gupta et al., 2004). Consumer free will to switch to better and/or less expensive choices in the retailing market means that it typically costs nothing for customers to switch from one retailer to another. Consequently, it is imperative for companies to focus on maintaining good customer relations and enhancing customer retention over the customer lifetime in order to generate higher profitability and growth in comparison to the significant costs and drawbacks associated with attracting and maintaining new customers (Aeron et al., 2008).

In customer retention efforts, segmentation is a critical tool for understanding how consumers differ in terms of their interactions with and behavioral responses to retail marketing (Wedel and Kamakura, 2002). Segmentation can be a powerful descriptor and drives more precise targeting and positioning, which can ultimately increase customer value (Foster et al., 2011). Using segmentation, firms may be able to offer products, services, and marketing campaigns that are very similar to those of other firms to many current consumers of those firms, who have a tendency to react or interact similarly. Time-tested segmentation approaches utilize information such as demographics, as well as psychographic methods, in order to surpass superficial customer segmentation and better comprehend purchase motivations and additional behaviors (Boone and Roehm, 2002; Malcolm and Dunbar, 2012; Wiertz and DeRuyter, 2007). Developing effective retail marketing methods first means understanding consumer buying behaviors.

The concept of brand personality recently has been analyzed extensively in marketing and consumer literature as a segmentation

* Corresponding author.
E-mail addresses: lukech@saturn.yzu.edu.tw (L.-L.L. Chiang), csyang@saturn.yzu.edu.tw (C.-S. Yang).

http://dx.doi.org/10.1016/j.techfore.2017.06.034
Received 22 January 2017; Received in revised form 21 June 2017; Accepted 23 June 2017

variable (Keller, 2001). The concept has been described as the human personality traits that a brand is associated with (Aaker, 1997). While an individual's personality may be displayed in behavior, beliefs, and physical traits, brand personality is capable of being formed via direct/indirect consumer brand experience (Kim et al., 2010). Thus, consumers can have personality traits similar to those of their preferred brand personalities. Brand personality traits can describe a brand in terms of a mutual cohort of the affiliation between brand and consumer (Sweeney and Brandon, 2006). Furthermore, brand personality has been found to have a positive effect on consumers' loyalty, satisfaction, and recommendations to peers (Keller and Richy, 2006; Lin, 2010). Having a remarkable “personality” makes a brand special in consumers' minds, thus forming and enlarging the equity of the brand (Freling and Forbes, 2005). However, while this has been noted by many scholars (Moilanen and Rainisto, 2008; Mulyanegara et al., 2009) brand personality has been applied infrequently to the context of national, country-of-origin (COO) brands in comparison to product or corporate brands.

Country-of-origin can be also perceived as a brand via the typical view held by consumers of the products/brands of a given nation (Bluemelhuber et al., 2007). As such, in a single advertisement two distinct brands (both the nation itself and the advertised brand/product) and their relationships with the consumer can be present. Country-of origin brand personality is currently salient to marketers because it is critical to differentiate nation brands (which are known to affect customer purchase decisions) (Yasin et al., 2007).

Country-of-origin brand personality is the association of individual personality characteristics with a COO product/brand. Positive COO brand personalities are typically preferred by customers and build trust, loyalty, and assurance. They also allow for different types of positioning strategies among the myriad marketing techniques in the competitive retail industry, thus potentially influencing consumers' purchase intentions (Dinnie, 2008). Marketers emphasizing COO brand personality should strive to match the personality or self-image of the consumer via segmentation-based efforts (Casidy et al., 2009; Lin, 2010). In addition, COO brand personality is an important component of consumers' emotional perceptions, which affect buyer perceptions and evaluations (Avis et al., 2012). In this regard, a COO brand personality approach can make valuable contributions to the development of an effective customer-driven marketing strategy.

Another stream of customer-based marketing that has received considerable attention involves customer lifetime value (CLV), which is the current valuation of projected profits over the course of a consumer's “lifetime” (relationship) with a company (Benoit and Van den Poel, 2009). Increasing numbers of marketers and performance evaluators have an eye toward relationships with customers, and retailers can draw substantial benefits from understanding more about CLV. As Kumar and Rajan (2009) stated, a critical challenge for companies is the implementation of the right “mix” of consumer treatment according to loyalty levels in order to maximize financial gains. A useful way to identify customer value/profitability of customers is the CLV concept, which employs recency/frequency/monetary analysis in order to develop segmented markets (Gupta et al., 2006). However, no standardized method exists to measure customer lifetime value, and the extant literature is limited with regard to actual application of the envisioned methodologies. Chan et al. (2010) emphasized that there are very few simple, robust, flexible, empirically valid CLV measurement models in the field of retailing, and there are few studies thus far that address CLV in a significantly better way.

A review of the previous research on CLV indicates that it is centrally concerned with customer personality traits as reflected in products and that it has been infrequently applied to the context of country-of-origin (COO) brands. Furthermore, guidelines on optimal marketing decisions for maximizing CLV and predicting potential CLV are scarce. Those studies that do address these issues in great part make educated guesses and lack models verified empirically; they are primarily based in theory alone (Haenlein et al., 2006; Jain and Singh, 2002). An important attribute of the present research is the use of state-of-the-art data mining techniques for examining COO brand personality and predicting potential CLV in order to create a comprehensive model that can be applied when making marketing decisions. More specifically, the findings of this study can help practitioners to gain an understanding of whether the COO brand personality should be used in targeting specific customer segments and the RFM model should be congruent to COO brand personality, and how these factors impact customer lifetime value in the retailing industry.

2. Literature review

2.1. Brand personality

Brands can possess attributes that mirror individual personality characteristics and are meaningful for both the brand and consumers (Azoulay and Kapferer, 2003). The personality of a brand is the way in which consumers define a brand based on its qualities and attributes (Parker, 2009). Furthermore, according to Plummer (1985), a brand can be defined as a mind-based creation developed by the consumer, who forms mental pictures about the brand through his or her experiences, perceptions, misconceptions, and value systems. This definition stresses that the creation of a brand personality is the result of an active and dynamic interaction of consumer/brand characteristics. Individual personality is revealed in the ways that people employ stable ways of thinking, feeling, and behaving (Keller and Richy, 2006. Both individuals and brands can have unique personalities. Given people's tendency to imbue “things” with human traits to facilitate interactions in an “intangible” realm (Sung and Tinkham, 2005), direct/indirect experiences with brands cause consumers to develop ideas about brand personality (Bosnjak et al., 2007). Lee and Cho (2009) stated that a brand personality is created from these experiences, and that perceptions do not take place without the influence of one's own needs, motives, and personality (Sweeney and Brandon, 2006).

In light of the ease with which competing companies can mimic product traits, creating and developing a distinct, cohesive brand is a difficult task for marketers. According to Mithas et al. (2013) successful branding means knowing how to build a brand personality that enables customers to view the distinctive personality of the brand and subsequently build a solid and enduring partnership with the brand. A successful brand personality is a real asset in differentiating a product from similar products. Thus, industry marketers can significantly benefit from developing and integrating features of brand personality in their marketing plans, using brand personality traits as an important guiding consideration (Sung and Tinkham, 2005). In order to be successful, a brand personality must have a very distinct differentiation that conveys and expresses that is unlike other brand personalities (Huang et al., 2012) and is carefully and continuously fostered over an extended period of time.

From a psychological perspective, consumers can develop relationship dyads with brands that are humanized by marketers (D'Astous and L'evesque, 2003). This relationship allows consumers to establish a reflexive evaluation of a product (Geuens et al., 2009). In this regard, brand personality strongly influences customers' brand choices, since customers typically chose brands that match their personal characteristics. A brand's matching “personality” traits provide avenues for consumers to symbolize and express themselves (Ahn et al., 2009). Specifically, Thomson et al. (2005) stated that consumers tend to select brands that match their real or ideal self-concepts or social self-concepts. Furthermore, brand personality may create genuine affection for particular brands, thus reinforcing consumers' views of themselves. In summary, brands have individual personalities, and consumers are more likely to choose products that match their own personalities and preferences (Romaniuk, 2008).

2.2. Country-of-origin (COO) as brand

The perception of a given nation strongly impacts how consumers perceive its products/brands (Ahmed et al., 2004) and influences buying behaviors (Hamzaoui-Essoussi et al., 2011). Jian and Guoqun (2007) found that the image of a particular country of origin (COO) can have a direct, positive effect on the evaluation of a product's quality by consumers. Han's (1989) study posited two functions of COO brand personality: the "summary effect" and the so-called "halo effect"; and proposed that customers unfamiliar with a given product/brand depend on "halo effects" that indirectly influence their attitudes during the process of inferring brand traits. In contrast, consumers familiar with a given brand or its products tend to sum up their attitudes toward them; this directly affects how they perceive them.

Furthermore, COO image may play a role in building nation brand equity and provide experiential and symbolic benefits to consumers during their purchase decision-making process (Hellgeson and Suphellen, 2004; Mulyanegara et al., 2009). For example, consumers may be attracted by nation-related personality traits such as French or Italian 'sophistication', Russian 'strength', Japanese 'serenity', and Mexican 'excitement'. Given that contemporary consumers tend to buy brands rather than products per se, the COO brand personality of products that rely on the traits of their COO to attract consumers should be an important consideration when developing marketing campaigns. Accordingly, Hamzaoui-Essoussi et al. (2011) explained that nation brand effects have shifted from product to brand level in consumer evaluations involving complex COO information. Consequently, these shifts have made the concept and personality of a brand more meaningful in terms of COO associations and have fueled new research by marketing scholars. Many scholars have suggested that brand origin is the geo-region or nation of the brand as viewed by target customers, and that cues pertaining to origin are embedded within brand personalities in relation to consumers' personality traits (Ahmed et al., 2004; Borle et al., 2008; Sanyal and Datta, 2011). Thakor and Kohli (1996) reconceptualized the COO concept (how consumers feel about a given country) and concluded that the brand origin concept (how consumers feel about the way that specific COO traits are incorporated into a brand personality) may be a more influential determinant of consumers' perceptions than COO. Nation brand personality can be summarized as consumer perceptions about the COO of a brand. In this vein, nation brand personality is an important element in building preferences for a particular brand because consumers link their wishes and aspirations to specific dimensions of brand personality. Thus, consumers use nation brand personality as a path to self-expression or to experience the emotional benefits by which a nation brand differentiates itself from other brands (Pappu et al., 2006). In addition, consumers may use a nation brand to express not only who they are but also who they would like to be or are expected to be by their social groups (Krohmer et al., 2007).

In particular, Aaker (1997) employed psychological theory pertaining to personality in order to construct a scale to measure brand personality along various dimensions of brand personality: excitement, sincerity, sophistication, competence, and ruggedness. In Phau and Lau's (2000) study, they developed a scale of brand personality using 36 traits by asking questionnaire participants to indicate their impressions. The general findings support the claim that consumers take nation brand concept into consideration when choosing products. However, research aimed at understanding and honing nation brand personality must consider the effects of consumer self-concept/personality, which is a crucial element in driving consumer behavior and thus in developing effective marketing strategies for retailing marketers (Pappu et al., 2006; Sanyal and Datta, 2011).

2.3. Advantages of data mining applications

Recently, data mining has been employed to gain a better understanding of the massive compilations of consumer data that many businesses compile daily (so-called "Big Data") in order to determine which customers are most prone to brand switching (Aggarwal, 2011). Numerous academics and industry practitioners have constructed various models that attempt to predict customers' future behaviors. In this regard, data-mining is less limited than many traditional, deterministic models because it can forecast these models' variables. Gelbrich and Nakhaeizadeh (2000) used prediction models in order to predict customer lifetime value directly. Donkers et al. (2007) sought to forecast an insurance company's customers' lifetime values using various models for mining data, while other researchers (Bolton et al., 2000) looked at loyalty program and service effects on consumer value/retention and created a model for demonstrating how loyalty rewards programs affect customer repurchase and frequency-of-use behaviors. The research body that strives to improve CLV prediction models continues to grow because of the importance of proactively targeting high-value customers using CLV prediction models via specifically tailored marketing actions. Furthermore, Gronroos (2007) stated that industry marketers can double their profits by taking precise and effective marketing actions toward their high-value customers using a predictive model, since maintaining current customers costs less than attaining new customers (who typically have a high rate of attrition) (Mulhern, 1999). Finally, CLV prediction research continues to be crucial since CLV prediction must be as accurate as possible because, as Ryals and Knox (2005) proposed, in maximizing the lifetime value or profitability of each customer, marketers also maximize the profitability and valuation of a business as a whole. Results to this end can be achieved by using applicable customer-based datasets and decision-support schemes.

For example, Kim et al. (2014) employed network analysis to examine a mobile telecommunications firm's subscriber communication patterns and were able to better predict churn (customer attrition). Specifically, they examined both a traditional machine-learning method for handling customer data and a propagation method that employed the call records of churning to non-churning customers. Heidemann et al. (2012) employed social network data in order to forecast consumer behavior and presented a new model integrating the effects of social networks in user-marketer interaction data using a relational learning algorithm (He et al., 2013). In their study, text-mining was utilized to develop measures to depict the persuasive and informative aspects of both user-generated and marketer-generated content and make a distinction regarding "directed" versus "undirected" communication. The significant impact of social media engagement on customers' potential behavioral intentions was discussed, and a prediction model was developed with the aim of increasing profits earned in a user-generated campaign.

In the recent past, data mining has been employed to develop solutions to challenging customer behaviors and CLV issues in the service field (Aeron et al., 2008; Kim et al., 2006). Due to the increasingly competitive nature of the service marketplace, data mining is used primarily to address customer behaviors or CLV prediction and has of late received unparalleled notice by service industry practitioners and researchers (Verbeke et al., 2011). It primarily involves the use of customers' log files in order to decipher the customers most apt to drive sales and to build and strengthen these relationships with a company, setting them apart from other customers in terms of value. Advanced analytics doesn't solely drive marketing decisions; data-driven insights can create value across the entire scope of a business. For example, cutting-edge retailers utilize data mining techniques to tailor product assortments at the store level, to minimize operations costs, and to anticipate changes in customer purchasing patterns, while simultaneously enhancing the customer experience and improving unit economics (Ngai et al., 2009). However, data mining can be highly expensive and time-consuming.

The majority of retailers today employ retailing information systems (i.e., point of sale, or POS, systems) in order to handle transaction and customer data, usually generating enormous quantities of data such as numerical figures, charts, text, and images. However, this data is infrequently utilized in developing marketing strategies in the retailing context, leaving a wealth of largely untapped information (Archak et al., 2011; Aggarwal, 2011; Verbeke et al., 2011). Thus it is a question of how to convert data to a practical format that allows retail professionals to create effective marketing and/or business decisions. The majority of the data mining techniques employed in retailing utilize various methods to attain more or less identical results, a valuable use of monetary resources for the majority of businesses.

2.4. Customer lifetime value (CLV)

Relationship marketing and the corresponding customer relationship management involve a radical change in approach in marketing management (Haenlein et al., 2006). Customer information databases and interactive, mass-customization technologies have led to customer relationship management (CRM) as way to understand and influence consumer behavior via targeted interaction aimed at improving customer acquisition, loyalty, retention, and profitability (Peppers and Rogers, 1999). Customer relationship management assumes that consumers have different needs as well as values that they generate for a business. As a result, CRM does not aim to offer all customers the same service; instead, it offers different levels of service based on customer lifetime value (Kim and Yoon, 2004); in other words, on the existing value of projected profits gained by a given customer across the "lifetime" with a firm (Gupta et al., 2006).

Various classifications exist for customer lifetime value models. Gupta et al.'s (2006) study offers five modeling methods, the first of which encompasses recency, frequency, and monetary (RFM) models. Next, they examine probability models adopted from the Pareto and NBD models; Markov chains; and econometric models such as the probability model adopted from the Pareto and NBD models and from customer acquisition, retention, margin, and expansion. They then examine persistence models adopted by modeling component behavior (acquisition, retention, and cross-selling). Finally, they explore models from computer science that emerged from sources such as utility theory and that offer ease of interpretation (Benoit and Van den Poel, 2009; Ryals and Knox, 2005). The vast extant literature in the social sciences on machine learning, data mining, and causal inference also resulted in diffusion and growth modeling adopted from customer equity (Berger and Bechwati, 2001; Reinartz and Kumar, 2003).

The recency, frequency, and monetary value model is a highly powerful yet simple way of implementing customer relationship management (Shih and Liu, 2008). The RFM model has been defined by Bult and Wansbeek (1995) as encompassing the factors of: (1) recency, defined as the amount of time since the most recent purchase, with lower values corresponding to higher probabilities of repeat purchases; (2) frequency, defined as the total purchases completed in a given time period, with higher frequencies indicating higher loyalty; and (3) monetary, defined as the amount of spending over a particular time period, with higher values indicating a need to focus greater attention on those customer groups. Considerable research has been conducted to propose and examine various approaches to using RFM models (Cheng and Chen, 2009). In recent research, some authors have proposed the use of the WRFM (weighted RFM) model, with different weights assigned to the recency, frequency, and monetary parameters, depending on the qualities of a given segment of industry (Chuang and Shen, 2008). For example, based on feedback from expert groups (i.e., administrative and sales managers, marketing consultants, etc.) and customers that made at least one prior purchase, Labbi and Berrospi (2007) suggested that weighting frequency, recency, and monetary from highest to lowest in that order could normalize the values. The RFM values of customers are strong predictors and such models offer convenience of use (Gupta et al., 2006). Thus, the study aims to create a successful model and exploit computer science tools to forecast CLV by accounting for the tool and the variables described both in the literature and by experts.

3. Methodology

3.1. Data collection and data mining structure

In the present study, data from 25,723 customers with approximately 44,000 transaction records from between 2013 and 2014 were obtained from the point of sale (POS) database of the city'super specialty store, a retailer in Hong Kong that offers food, beverages, and lifestyle products. Specifically, the data were examined with regard to beer purchases. In the Big Data era, it is critical to gain an understanding of the heterogeneity of consumers. Big Data can be employed to understand heterogeneous groups and their statistical properties, and to extract important signals of the presence of large individual variations. Therefore, it is conceived of as a data mix that arises from heterogeneous variations. Accordingly, our data collection process accommodates customers' varying purchase behaviors and considers customers neither on the individual level nor as a whole. Instead, customers are segmented according to various criteria, allowing for the identification of groups of customers that share similar purchase behaviors. For instance, some customers may be particularly attracted to German or British brands because these COO brand personalities reflect their own personality traits (i.e., peacefulness, sincerity, and excitement, etc.).

Data-mining is multi-step, iterative process. This study employs the tool-neutral CRISP-DM methodology used in the industry (e.g. Daimler Chrysler, SPSS), in order to build the mining models. The six phases of the dating-mining process are as follows:

1. Business understanding
2. Data understanding
3. Data preparation
4. Modeling
5. Evaluation
6. Deployment

The first phase includes a retail business perspective assessment of the requirements/objectives in order to develop a defined problem and design an initial program to attain the stated goals using data mining. The first phase employs unrefined data from customer transaction datasets in order to attain initial insights into the data and find salient data subsets. The second phase involves the analysis of unrefined data (the customer transaction dataset) in order to attain initial insights and find salient subsets. In the third phase, the final dataset that will be fed into the modeling tools (i.e., R tool) is constructed. This phase includes the selection of tables, records, and attributes, and data cleansing and transformation. The fourth phase involves the selection and application of a variety of methods and the calibration of relevant parameters to the values that have been determined optimal. In the fifth phase, model evaluation is carried out to verify its ability to achieve the stated objectives. In the final phase, the required steps for using the models are specified. In the present study, steps 4 and 5 emphasize the use of network and associative analyses and multivariate methods to achieve classification and occurrence goals among variables (i.e., beer brands, beer types, country of origin, customer personality traits, etc.) where perceptual graphics of a model can offer information that illuminates classifier performance. This paper follows the CRISP-DM methodology structure (see Fig. 1).

3.2. Network and co-occurrences analyses

In this paper, two important and common approaches were proposed to develop a highly inclusive view for an enriched comprehension of consumers' purchase behaviors, taking into account customer and COO personality traits to predict potential CLV. First, data were

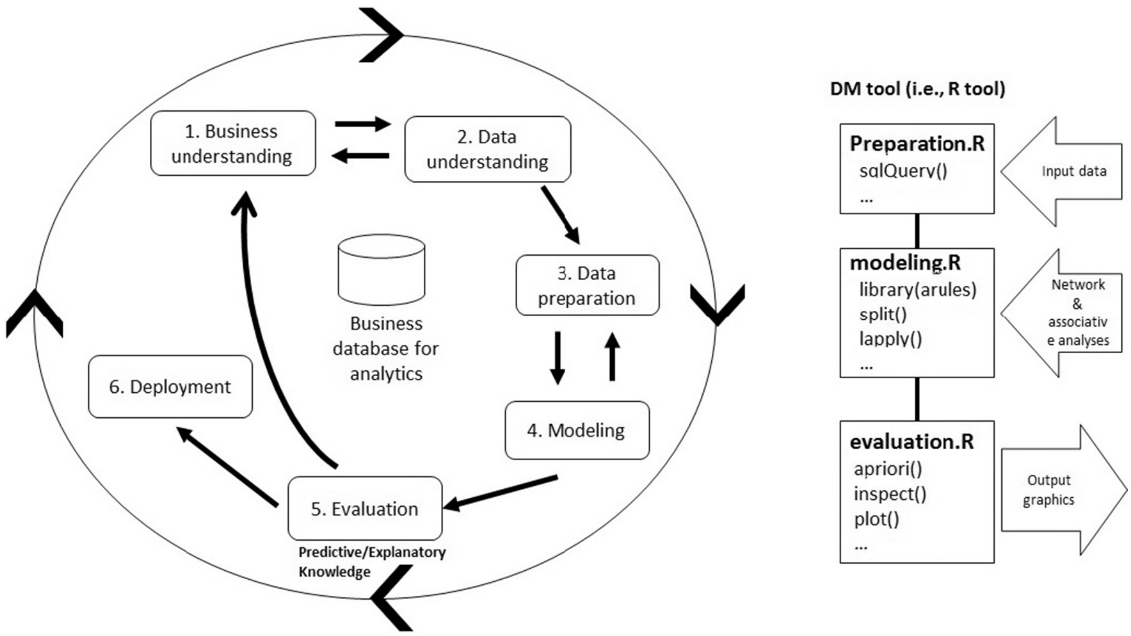

Fig. 1. The CRISP-DM and proposed DM tool to use.

analyzed using the R tool, which is a multi-platform (e.g. Windows, Linux, Mac OS), open-source, high-level matrix programming language employed in the analysis of statistics and data. While it is not specially developed for data mining or business intelligence, it offers a significant array of dating mining algorithms and currently is utilized by many data mining and business intelligence analysts (Cortez et al., 2009; Williams, 2009). When compared with commercial tools, the R tool offers greater flexibility and is more extensible, permitting the more natural incorporation of programming, statistics, and graphics. In light of these factors, the present study posits that the R tool is a worldwide gateway for sharing computation algorithms for further analysis.

Second, the multivariate methods and network analyses employed in this study for transaction data analysis were employed to create perceptual maps depicting the similarities and dissimilarities between beer purchases that emerged from the city'super dataset. Specifically, an investigation was conducted regarding the number of co-purchased beers and the terms in the transaction data pertaining to co-occurrence patterns seen in the extracted purchasing network. Next, co-occurrences were analyzed to determine beer purchase patterns in the transaction data and create perceptual mapping of purchasing networks and market structures. In order to identify the optimal performance for classification and co-occurrence purposes, the lift measure (originally called Interest) was also conducted among the 74 beer brands using the beers' characteristics (brands, types, and COO), independent purchases of the beers in the transaction records, and beer purchase co-occurrence with terms that consumers most commonly used to purchase beer at city'super. The lift measure assesses the number of times A and B (i.e., brands, types, and COO) occur together compared to the expected number of times if they were statistically independent, and was employed in the present study due to its intuitiveness when measuring co-occurrence against chance expectations (Turney and Littman, 2003). Lift measure can capture the co-occurrence and thus similarity between pairs of terms and is defined as the following equation:

$$Y = P(A, B)/P(A) \times P(B)$$

where Y = lift (A,B), is the lift between terms, P(A) is the probability of occurrence of term A in a given transaction record, and P(A)(B) is the probability that both A and B appear in a transaction record. Accordingly, the detailed lift definition among variables in this study are presented as:

Brands/types:

If Y = 1, then A and B appear as frequently together as expected under the assumption of conditional independence. A and B are said to be independent of each other (e.g., *Asahi* Super Dry/*Asahi* Style Free or *Anchor* Steam and *Anchor* Liberty); and 0 otherwise.

Country of origin:

If Y = 1, then A and B appear as frequently together as expected under the assumption of conditional independence. A and B are said to be independent of each other (e.g., *Belgium* Lindemans and *Belgium* Trappistes Rochefort); and 0 otherwise.

3.3. Customer lifetime value and RFM analyses

Two clustering methods were used to segment customers for assessing customer lifetime values. In the first approach, COO brand personality and co-occurrence parameters were included in clustering. Like image segmentation and web mining research, cluster analysis (CA) is one of the most popular of the many DM techniques that can be employed in consumer data analysis. Research has urged the use of DM and CA because they can improve the accuracy of marketing decisions (Brengman et al., 2005). Among the various clustering techniques, K-means (KM, or Forgy's method) is a well-known clustering algorithm that has been broadly utilized in statistical analysis, data mining, and additional business-related applications (Forgy, 1965; Ghosh and Dubey, 2013). By contrast, soft partitioning or fuzzy clustering algorithms, like Fuzzy C-means (FCM), is an extension of KM (Bezdek, 1981). Some scholars (Jipkate and Gohokar, 2012; Suganya and Shanthi, 2012) have posited that FCM is more efficient for analyzing fuzzy data; however, according to the research findings, it lacks a constant superiority in all cases of data structures. Many studies have reported that, while FCM offers superior results for noisy, clustered datasets (Bora and Gupta, 2014; Ghosh and Dubey, 2013; Kaur and Kaur, 2013), KM is the preferred choice for large datasets because of its relative efficiency in computational time complexity. Thus, the use of KM in this study was a good starting point for large customer transaction datasets due to its fast execution time. In order to identify optimal *k* (e.g., how many clusters each individual clustering is comprised of), various metrics can be employed. Accordingly, the number of optimum clusters can be obtained with the value of recency, frequency, and monetary and by calculating customer lifetime value for each cluster along with the weighted RFM, the approach employed in this study.

4. Empirical analysis and results

4.1. Co-occurrence and COO beer brand personalities

The DM literature commonly summarizes consumer future behavior and intentions based on past behavior using data from extensive databases of customer-driven transactional data, figures, and socio-demographic data. The dataset in this study involved a robust and popular category comprised of many products and multiple product COO brands, with a plentitude of data from enthusiast and non-enthusiast customers. The data-mining process described in the Methodology section was employed to convert a quantitative and structured dataset consisting of thousands of transaction data into a purchasing network of co-occurrence of beer brands in the city'super database. This method offered the advantage of presenting structured data graphically. Certain network subsets/domains can then be monitored and the relationship between information can be traced in greater detail. In this study, a 74 × 74 lift dyad matrix between pairs of beer brands and types stored in the dataset was constructed. Every node pairing had a correspondence with symmetrical strength as shown in lift magnitude: lift (i,j) = lift (j,i). The characteristics of the network were examined to achieve a better understanding of the nature of the dataset and the centrality of different beer brands and types to customer purchasing patterns and customer personalities. Despite the large number of beer brands and types involved, the dyad matrix was relatively dense (67.8%) owing to data richness (nearly 44,000 transactions). That is, on average, 67.8% of the beer type pairs were co-purchased at least once. Network centrality measures (the centrality of "betweenness" and eigenvector) were also utilized to assess the relevance of various beers in the transaction.

A figure was developed (see Fig. 2) to offer a visualization of the network of purchase of the 74 beer brands with types in the dataset by employing a spring-embedded algorithm (Kamada and Kawai, 1989) that minimized system stress in the connection between network nodes such that more similar (higher lift) beer brands and types were closer together. The edge width that connected the nodes (beer types) displayed the lift magnitude of beer types. Lifts significantly superceding 1 at the 99% level according to the results of the χ^2 test (see Fig. 2) are dense yet highlight the advantageous aspects of employing both structured mining of data and analysis of network methods to follow customer transactions and co-purchasing patterns. Data mining was used to simultaneously measure the transactions for 74 beer types with much less difficulty and cost as compared to typical marketing research methodologies. Next, network analysis allowed for the analysis and visualization of a significant variety of entities in order to offer a thorough portrayal of the transactional dataset.

Based on their personality traits, consumers tend to choose and purchase beers with brand personalities that are similar to their individual lifestyle characteristics. In this study, five adjectival descriptors of personality traits were used for further analysis. The results shown in Fig. 2 (with beers belonging to the same family grouped together in a variety of colors) provided high face validity. For instance, the cluster of beers indicated by the color orange includes, for the most part, the "innovation" dimension (e.g., beers from Belgium, Germany, and the UK). In addition, the innovation dimension taps into traits of openness. The cluster of beers represented by the color red includes beers that fall primarily under the category of Japanese beers (e.g., Asahi, Suntory, etc.). The cluster of beers designated by red squares includes beers that belong predominantly to the "sincerity" dimension of traits. The colors of the 74 nodes in Fig. 2 represent dissimilarities of membership, a finding that is discussed in greater detail later in the study.

To further explore the role of customer personality traits in customers' perceptions of COO brands and types, the study sought data regarding clusters of beer types purchased together frequently. The Girvan and Newman (2002) community algorithm (often employed in network clustering) was used, resulting in clusters consisting of node groups connected densely in clusters and less densely across clusters. In contrast to traditional network analysis, the purchasing network employed in the present study consists of 'communities' of beers rather than of people. According to the results obtained by the use of the Girvan–Newman algorithm, this study streamlined (or trimmed) and identified 40 nodes of beer brands. While this number seems large, only five large clusters emerged. For a clearer image of the associative network and perceived market structure, the study combined the purchasing co-occurrence of beer brands. In total, the city'super dataset showed the purchase of 40 beer brands. As with beer brands and types analysis, lift between brands purchased together was used to assess similarity and association. Due to the manageable quantity of brands, it was possible to use multidimensional scaling (MDS), a traditional tool to visualize and analyze market structure (see Fig. 3). The coordinates of MDS were utilized to perform cluster analysis on the derived MDS, with dashed ovals reflecting the five-cluster solution.

Next, adjectives describing COO brand and customer personalities were analyzed in terms of whether they exhibited congruency with the Smith et al. (2006) BP scale and to determine the extent to which these

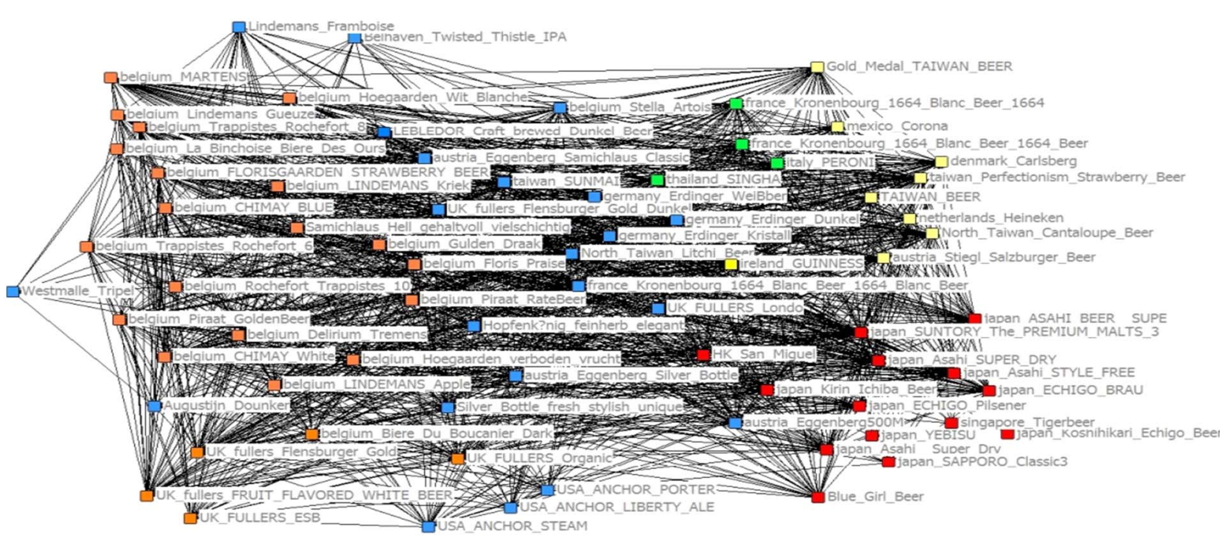

Fig. 2. Network and associative graph of beer brands and types.

Fig. 3. MDS map of COO brand and customer personality traits.

adjectives were clustered around a specific dimension of the scale (Smith et al., 2006). The findings served to explore personality perceptions toward the COO brands of beers and to determine the major dimensions of the current BP scales to best present these perceptions. Several insights can be derived from the results shown in Fig. 3. First, the majority of brands from the UK, Germany, Belgium, Austria, and the USA were clustered at the bottom right of the map. Customers can typically characterize themselves in terms of personality trait adjectives such as 'openness', 'new experiences', and 'innovativeness' in this cluster, and customer profile attributes are associated with creativity and preference for novelty and variety. In addition, the trait of openness was defined empirically as possessing varied interests and atypical thoughts (indicating original and unique behavior) and unconventional judgments. This implies that individuals with high openness levels may inquisitively focus on a variety of product aspects rather than a single aspect such as product appearance.

In the middle left part of the map are local brands such as Taiwan Beer, Taiwan Labledor, and Taiwan North Beer. This result implies a COO brand personality of sincerity and genuineness. Customer profiles of this cluster tend to be kind, family-oriented, and thoughtful and such consumers associate with peers and friends that share the traits of wholesomeness, genuineness, and honesty. The factor of sincerity is comprised of four parts: honesty, being down-to-earth, wholesomeness, and cheerfulness. In this regard, Taiwan is a relational society in which interpersonal bonds play an important role in building and maintaining social relationships. Therefore, customer personalities in this cluster expect relationships to have strong emotional content.

In the bottom left part of the map are multicultural brands and mainstream Japanese brands such as H.K. San Miguel, Singapore Tiger, Japan Kirin, and Japan Suntory, etc. This finding suggests that customers who belong to this cluster perceived relational attributes pertaining to the trait of 'peacefulness', perceived as the naivety and mildness of a brand personality. Customer traits in this cluster are associated with a unique blend of attributes such as peacefulness, shyness, naïveté, and dependence, which closely reflect the indigenous peoples of these countries.

A number of South American and some European brands (i.e., Mexico Corona, Denmark Carlsberg, Netherlands Heineken, Ireland, and France Kronenbourg, etc.) are found in the top and middle sections of the map. The cluster was positioned on the traits of the 'excitement' dimension. Excitement is the amount of freedom, talkativeness, energy, and happiness in brand personality, and is related to a sense of being carefree and spirited, with the four factors of daring, spiritedness, imaginativeness, and being up-to-date. Youthful and 'cool' customers are associated with this factor.

It was found that Italian and French beer brands were purchased more frequently with the Thailand Singha brand. In this case, one dimension found and labeled on the cluster is 'sophistication', the style and elegance of brand personality that involves aspirational images associated with wealth and status. The 'sophistication' dimension of customer personalities has associations of upper class appeal, with traits of charm, glamour, and smoothness. More specifically, customer profile attributes for these brands are more likely to be elegance, prestigiousness, and pretentiousness. These results are consistent with the results found by Smith et al. (2006) and Aaker et al. (2001).

4.2. Estimating CLV performance for customer clusters

In order to develop a more comprehensive, customer-focused model to guide retailers in making future marketing decisions, customer lifetime value analysis was used in this study. For the five clusters of segmentation, the average CLV was calculated and each segment was assigned a CLV ranking. To calculate CLV for the five clusters, the weighted RFM method was used. According to retailing experts' suggestions, RFM variable relative weights are: $WM = 0.437$, $WF = 0.366$, and $WR = 0.197$. The weighted value of M is the highest to represent revenue gained from a customer, which is the most important factor in the retail industry. The weighted value of F (the number of individual customer purchases) is the second most important and can be used to represent customer loyalty. The R value, consisting of the most recent customer purchase, has the lowest weighting. In addition, R, F, and M are non-homogeneous parameters; therefore, these parameters need to be normalized (Shih and Liu, 2008). Based on the weighted RFM method, the largest average CLV value is 0.3148839 (peacefulness group), followed by 0.2635248 (openness group), 0.1765119 (excitement group), 0.1248102 (sincerity group), and 0.0659046 (sophistication group).

In order to analyze the clusters, it was necessary to place the recency, frequency, and monetary parameters into the distinct categories of Very High, High, Medium, Low, and Very Low, as discerned by retail

Table 1
RFM categories.

	Very low	Low	Medium	High	Very high
Recency	1–145	145–290	290–435	435–580	580–725
Frequency	52–99.8	99.8–147.6	147.6–195.4	195.4–243.2	243.2–291
Monetary	7600–263,520	263,520–503,468	503,468–759,388	759,388–1,015,308	1,015,308–1,271,288

Table 2
RFM analysis and CLV ranking for each cluster.

Cluster	R	F	M	CLV value	CLV rank
Excitement	High	Medium	High	0.1765119	3
Sincerity	Medium	Medium	High	0.1248102	4
Sophistication	Very low	Low	Medium	0.0659046	5
Peacefulness	High	Very high	Very high	0.3148839	1
Openness	High	High	Very high	0.2635248	2

Table 4
The performance measures for the extended customer pyramid.

Segment	Hit ratio
Top 20%	0.6481
Next 30%	0.5680
Latter 30%	0.5923
Bottom 20%	0.7258

industry experts (see Table 1). Table 1 shows the comparison of the CLV values and results of RFM parameter values (with categorical values) by cluster; a summary of the category of the CLV rank for each cluster is provided (see Table 2).

Next, the performance of another CLV prediction model was assessed using the hit ratio that accurately predicted customer segments and equals how many customers the model predicted to fall within a given segment divided by the actual number of customers in the segment. The proposed model that employed the 2013 and 2014 datasets in the study was used to predict potential customer values, and predictive accuracy was assessed using criterion of the hit ratio (Donkers et al., 2007; Malthouse and Blattberg, 2005; Zeithaml et al., 2001). The top 20% of customers forecasted as compared to the actual top 5145 customers was calculated (see Table 3) and was found to be close to a classifying table in Malthouse and Blattberg's (2005) study. The table reveals that, of the 20% most profitable customers (top 20%) in 2013 and 2014, 35% could not be identified by using the model. Likewise, it was found that the 12% least profitable customers (bottom 80%) had been classified erroneously. The 20–55 and 80–15 rules derived by Malthouse and Blattberg (2005) were used to clarify these results. These rules provide an evaluation of a business's accuracy in estimating future customer value via several datasets from distinct industries. Based on Malthouse and Blattberg's estimations, it was discovered that approximately 55% of the actual top 20% experience misclassification (the 20–55 rule) while approximately 15% of the actual bottom 80% experience misclassification (the 80–15 rule). In light of the above-cited study (2005), the present study provided highly competitive results.

In order to develop an empirically and widely applicable model to predict CLV in the context of retailing, the study also divided the customer pyramid into four additional segments, namely 20%, 30%, 30%, and 20% of customers. Performance measures for this different method were also calculated. The poorest hit ratio was 56.8% for the second segment (see Table 4), an acceptable result in light of the 20–55 and 80–15 rules. The 56.8% hit ratio indicated that the proposed model (see Table 4) achieved correct classification of approximately 56.8 of 100 customers for the upcoming year. The model very accurately forecasted the most and least profitable customers. Table 4 shows that the average value of the first group is very high and maintaining these relationships is vital for two reasons. First, high-value customers will return to companies even if they can find similar products at lower prices. It's not about the cost: it's about how the business addresses their specific needs in a relevant and personalized way. Second, many high-value customers are influencers who may not spend a lot with the company regularly but provide incredible promotional outreach, often via social channels. In fact, the groups of the top 20% and bottom 20% of customers display more stable behaviors and likely were in identical segments in 2013 and 2014 consecutively. It is believed that other influences may have affected the other two groups between the extremes and that they are potentially unstable in terms of profitability; thus, their profitability is more challenging to forecast.

Table 3
Classification of the actual and predicted 20–80% groups for future CLV.

	Actual		Total
	Bottom 80%	Top 20%	
Predicted			
Bottom 80%			
Number of customers	18,116	2462	20,578
Hit ratio	0.88	0.12	0.8
Top 20%			
Number of customers	2462	2683	5145
Hit ratio	0.35	0.65	0.2
Total			
Number of customers	20,578	5145	25,723

5. Discussion and implications

5.1. Theoretical implications

This work most importantly contributes as the first study to empirically classify personality-based consumer perceptions of goods with nation brands by exploring and analyzing vast amounts of transaction data. Although there are segmentation and clustering methods for a variety of online consumer behavior and customer relationship typologies (Aggarwal, 2011; O'Loughlin and Szmigin, 2006) (Rohm and Swaminathan, 2004) and the literature presents a number of brand typologies, there is a dearth of segmentation or clustering methods that assess brand personality traits via data mining. The present study provides new, empirical insights into consumer self-personality and nation brand personalities by applying a number of Big Data analytics. The essential importance of these findings is the new ability to classify the many “real world” nation brand personalities through nation brand personality and consumer self-representation clusters and provide a jumping off point for further research. In addition, the value of the segmentation approach lies strongly in its ability to build upon dimensions adopted from the extant literature (i.e. Aaker, 1997) and provide a conceptual model, in contrast to the existing research (Malcolm and Dunbar, 2012.

The research also offers the provision of greater insight into converting numeral transactions to the performance aspects of the market structure. The authors utilized data mining to supersede the challenges inherent in extracting and quantifying the extensive retailing information systems data generated from consumer transactions, and employed network analysis tools to convert co-occurrence relationships into visualization maps among brands or between brands and terms (i.e., types). The resulting network with multidimensional scaling obtained from the dataset offers a high degree of density. This investigation of nation brand personalities and consumer personalities that drive the co-occurrence of beer purchase data reveals that consumers belong to clusters of similar personality traits and/or behaviors with respect to adjectives that are associated with the beers themselves (i.e., sophistication, peacefulness, excitement, etc.), the traits that consumers receive from purchasing the beers, and the beers' nation brand personalities. It also reveals that consumers are more likely to co-purchase beers with similar brand personality traits. These association analyses within brand personalities reveal how data mining permits zooming into data in a POS system in order to examine the competitive market structure via the consumer tendency to gravitate toward particular product attributes (i.e., beer brands and types, etc.). Thus, businesses can employ the data-mining process detailed in the present study and monitor their market positions longitudinally at higher resolution with greater effectiveness and efficiency compared to the use of traditional data sources such as survey-based consideration datasets.

Third, the ability to reveal potential consumer lifetime values achieved in this study is especially useful for facilitating competitive capabilities in light of the growing importance of customer relationship management. In addition, this study takes into account the vital significance of customer lifetime value and closes the literature gaps regarding customer lifetime value modeling. By using extended consumer-related parameters (i.e., nation brand personality and/or consumer self-personality), the RFM model, and the CLV paradigm, the major contribution of this research is the development of an applicable and industry-specific model with five different clusters of customers that offers a map for making marketing decisions based on projected CLV.

Finally, the study developed three weighted recency/frequency/monetary factors (namely WF-0.366, WM-0.437, and WR-0.197) through in-depth interviews with industry experts in order to create an industry-specific model. The proposed model confirms that the common objective parameters covered in the literature are significant and valid. Furthermore, the results indicate that five different nation brand personality and consumer self-representation groups (excitement, sincerity, sophistication, peacefulness, and openness) in certain retailing services are additional important parameters to measure potential customer lifetime value. Since the previous literature has been limited to retailing industries that often employ survey-based methodologies that do not uncover hidden consumer insights, the present study offers improved methods to determine potential customer value and other consumer-related factors through the use of Big Data that can be employed to develop marketing activities and sustainable competitive advantages for retailing businesses.

5.2. Managerial implications

Because marketers in the retail industry continuously seek methods for improving campaign effectiveness, targeting consumers with specialized offers designed to induce repeat purchase behaviors is a highly useful strategy. Identifying the traits of a target segment allows for the development of appropriate brand positioning strategies and enables marketing practitioners to create promotional themes that incorporate brand personality traits that match those of the target consumer. The ultimate marketing goal is to make the most relevant match between customer and offer to generate higher profitability and loyalty. In this regard, Big Data is the new capital in today's competitive service industry marketplace (Lycett, 2013; Rijmenam, 2014). However, as discussed in the study, the conversion of Big Data into a sustainable competitive advantage is a complex process. The present study offers a uniquely effective means to gain this advantage via the use of datasets that detail shopping behavior to extract consumer insights, as well as to make use of insights within the data to improve adaptive capabilities.

In terms of marketing, personality is a key aspect in influencing buyer behavior. For example, if an individual is more concerned about their id, are less ego-driven, and have a weaker superego controlling his or her behavior, they are more likely to make less rational decisions. This consumer is more inclined to make purchases simply out of opportunity presenting itself, without taking into account his or her financial situation or the effects of their actions on others. Furthermore, one's personality can also be related to the specific type of brand and product that he or she purchases. This study demonstrates that five key competitive characteristics of sixteen nations can reflect consumers' personalities and/or self-representations. For example, sincerity is a trait associated with the nation of Taiwan. Consumer personality traits in the "sincerity" cluster reflect sentimentality, warmness, genuineness, and wholesomeness. Taiwanese individuals thus may express themselves in part by the "sincerity" dimension of the goods that they buy, especially when the brand personality is socially visible among peers. For some, purchasing a Taiwanese beer expresses a sentimental self, based in part on the perception that the Taiwan-brand-as-person is unpretentious and warm. Knowing this enables marketers to have a deeper understanding about the nature of the relationship between the dimension of sincerity and its relevant customers.

It is important to identify a nation brand's specific competitiveness in terms of COO personality in order to entice consumers, who are more likely to select products that match personal preferences and traits. Nation brand personality encompasses a wider potential range of features and traits for a product or corporation, and therefore requires a more sophisticated managerial approach from the retailing perspective in order to maximize cohesiveness with consumer personality type. This endeavor can be advantageous from many perspectives (Dinnie, 2008; Moilanen and Rainisto, 2008) when seeking to attract consumers and promote the goods and services that contribute to the development of an effective customer-driven marketing strategy. Therefore, identifying the most salient nation brand personality traits in relation to those of the consumer plays a critical role in attracting varied target audiences through carefully designed retail marketing strategies. A highly distinct nation brand personality and concept lead to higher evaluations of nation brand personality attractiveness when they match the high self-expressive values and/or personality traits of the target consumer.

In addition, nation brand personality factors were found to be significant in predicting consumers' intentions to recommend a brand or product. The dimension of "excitement", for instance, was dominant as a nation brand personality trait for Mexico, the Netherlands, Denmark, and Ireland, appealing to consumers drawn to a sense of freedom, imaginativeness, being up-to-date, and energy. More specifically, the traits of youthfulness and "coolness" in customers are associated with the "excitement" factor, and retailers can imbue high-excitement nation brands' marketing campaigns with greater appeal by adding interest and involvement for this group of consumers to effectively amplify consumers' perceptions and experiences.

As a practical matter, marketers in the retailing industry should seek to implement two-way, brand-consumer strategies that match the dimensions of nation brand personality (i.e., "peacefulness", "excitement", etc.). These strategies may include advertising, physical packaging, promotions, events, customer touch points, digital programs, and more. Moreover, once the targeted consumer segment traits are notated and the position approach for nation brand is conceived, retailing marketers must create appropriate avenues for promotion that link the nation brand and the consumer in terms of personality. If marketers clearly perceive the variety of consumer types, they should then develop related campaigns to segments with the highest interest levels. Specifically, marketers should seek to develop messages that resonate with target consumers' most relevant personality characteristics as points of reference for guiding behaviors. These overarching themes are instrumental in creating successful promotional strategies such as online ads, packaging, public relations, and events.

The extended theory of brand personality and the positioning of a number of nations in terms of consumer self-representation that have been generally examined in this study, in combination with results pertaining to CLV prediction, can be used as an important guideline for creating highly precise sales and marketing strategies. Since customer lifetime analysis typically seeks to identify the customers that most significantly drive profits (the 20% that generate 80% of profits) of a given product/brand, customers were grouped in the model into the top 20% and bottom 80% in terms of hit ratio. Furthermore, it is obvious that marketers in the retailing sector need to prioritize the use of certain traits and personality themes in their goods and services, as well as to persuade targeted high-value customers (for example, customers with traits associated with "peacefulness" and "openness" might be a desired target segment) to buy those goods more frequently. Specifically, up-to-date, customer-driven strategies are highly useful for retailers seeking to retain and extend CLV (Gupta et al., 2006). The results of this study also show industry practitioners the importance of storing customer data for further use, mining the data with specific objectives, and using optimization techniques to make decisions about future marketing plans. In particular, customer features or traits-centric expansion is critical to driving competitive differentiation and increasing CLV.

Finally, the strategies suggested in this study can influence the customer loyalty, retention, and advocacy that drive satisfaction and result in increased growth and revenues. These issues are particularly relevant in the contemporary marketplace due to the reduction in entry barriers fostered by the Internet, resulting in widespread competition across the global landscape. Customer lifetime value does not pertain merely to maximized profits; it encompasses the maximization of profits from customers across the most extended duration possible. A focus on maintaining high-value customer relations, long-term maintenance of customers, and an overall reduction in customer churn can reduce the long-term costs per customer of a retail firm. Improvements in efficiency and cost reductions during customer acquisition, as well as increased growth and profits pertaining to loyalty, retention, and advocacy, typically result in increased customer lifetime values. Keeping high-value customers; satisfying them by delivering goods that strongly match customer self-personality; the use of up-selling and cross-selling strategies; developing custom promotional campaigns by segment; and increasing customer lifetime value are critical to the extended robustness and profitability of a company in the retailing industry over time. In summary, the data-mining methods, derived market-structure, and CLV analyses featured presented in the present study provide an entrée to explore the vast, rich, and valuable wealth of consumer data that is easily available in the retailing industry.

6. Limitations and future research directions

The present study possesses several inherent limitations. First, it utilized a quantitative transaction dataset from a single retailer located in Taiwan. It is recommended that additional datasets from different retailers and/or a larger array of countries be employed to attain wider comprehension of consumer behavior. Second, the study employed several industry-specific parameters in the CLV model, whereas the use of additional industry- and consumer-specific indicators such as those related to competition, customer satisfaction, and customer loyalty would be useful in future research. Specifically, these factors are critical in developing successful marketing strategies that offer customized service to targeted consumers, yet the necessary longitudinal and other data for various factors typically are procured through cross-sectional methods and thus lack integration into existing transaction databases. Third, future research could extend the application of multidimensional scaling and/or associative analyses beyond quantitative and structured data consumer transactions to include text mining and unstructured apparatuses such as blogs, online reviews, and formal news articles to allow businesses to better understand recent discussion online, the competitive landscape, marketing opportunities, and proprietary/competitor product features discussed by existing or potential consumers. Fourth, the study employed only K-means for clustering, thereby limiting the generalizability of the findings. Therefore, future studies should aim to also utilize Fuzzy C-means (FCM) for its benefits in terms of computing performance and clustering to generalize the findings and provide different insights. Last but not least, this study utilizes closed-source as opposed to open-source data. Future researchers should examine alternative datasets, the results of which can then be compared to attain improved access to consumer Big Data.

References

Aaker, J.L., 1997. Dimensions of brand personality. J. Mark. Res. 34 (3), 347–356.
Aaker, J.I., Veronica, B., Jordi, G., 2001. Consumption symbols as carriers of culture: a study of Japanese and Spanish brand personality constructs. J. Pers. Soc. Psychol. 81 (3), 492–508.
Aeron, H., Bhaskar, T., Sundararajan, R., Kumar, A., Moorthy, J., 2008. A metric for customer lifetime value of credit card customers. J. Database Mark. Cust. Strateg. Manag. 15, 153–168.
Aggarwal, C.C., 2011. An introduction to social network data analytics. In: Aggarwal, C.C. (Ed.), Social Network Data Analytics. Springer, United States.
Ahmed, Z.U., Johanson, J.P., Boon, L.C., 2004. Does country-of-origin matter for low-involvement products. Int. Mark. Rev. 21 (1), 102–120.
Ahn, K., Lee, J., Jeon, J., 2009. The effects of luxury brand-self identification on brand attachment and brand commitment: the moderating role of regulatory focus. Korean Mark. J. 19 (4), 1–34.
Archak, N., Ghose, A., Ipeirotis, P.G., 2011. Deriving the pricing power of product features by mining consumer reviews. Manag. Sci. 57 (8), 1485–1509.
Avis, M., Aitken, R., Ferguson, S., 2012. Brand relationship and personality theory: metaphor or consumer perceptual reality. Mark. Theory 12 (3), 311–331.
Azoulay, A., Kapferer, J.N., 2003. Do brand personality scales really measure brand personality. Brand Manag. 11 (2), 143–155.
Benoit, D.F., Van den Poel, D., 2009. Benefits of quantile regression for the analysis of customer lifetime value in a contractual setting: an application in financial services. Expert Syst. Appl. 36, 10475–10484.
Berger, P.D., Bechwati, N.N., 2001. The allocation of promotion budget to maximize customer equity. Omega 29 (1), 49–61.
Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York.
Bharadwaj, A., El Sawy, O.A., Pavlou, P.A., Venkatraman, N., 2013. Digital business strategy: toward a next generation of insights. Manag. Inf. Syst. 37 (2), 471–482.
Bluemelhuber, C., Carter, L.L., Lambe, C.J., 2007. Extending the view of brand alliance effects. Int. Mark. Rev. 24 (4), 427–443.
Bolton, R.N., Kannan, P.K., Bramlett, M.D., 2000. Implications of loyalty program membership and service experiences for customer retention and value. Acad. Mark. Sci. J. 28 (1), 95.
Boone, D.S., Roehm, M., 2002. Retail segmentation using artificial neural networks. Int. J. Res. Mark. 19 (3), 287–301.
Bora, D.J., Gupta, A.K., 2014. A comparative study between fuzzy clustering algorithm and hard clustering algorithm. Int. J. Comput. Trends Technol. 10 (2), 108–113.
Borle, S., Singh, S.S., Jain, D.C., 2008. Customer lifetime value measurement. Manag. Sci. 54 (1), 100–112.
Bosnjak, M., Bochmann, V., Hufschmidt, T., 2007. Dimensions of brand personality attributions: a person-centric approach in the German cultural context. Soc. Behav. Personal. Int. J. 35 (3), 303–316.
Brengman, M., Geuens, M., Weijters, B., Smith, S.M., Swinyard, W.R., 2005. Segmenting Internet Shoppers Based on Their Web-usage-related Lifestyle: A Cross-cultural Validation. J. Bus. Res. 58 (1), 79–88.
Bult, J.R., Wansbeek, T.J., 1995. Optimal selection for direct mail. Mark. Sci. 14 (4), 378–394.
Casidy, R., Tsarenko, Y., Anderson, A., 2009. Towards particular brand personality. J. Brand Manag. 16 (4), 234–247.
Chan, S.L., Ip, W.H., Cho, V., 2010. A model for predicting customer value from perspectives of product attractiveness and marketing strategy. Expert Syst. Appl. 37, 1207–1215.
Cheng, C.-H., Chen, Y.S., 2009. Classifying the segmentation of customer value via RFM model and RS theory. Expert Syst. Appl. 36, 4176–4184.
Chuang, H., Shen, C.C., 2008. A study on the application of data mining techniques to enhance customer lifetime value based on the department store industry. In: The Seventh International Conference on Machine Learning and Cybernetics. Vol. 1(3). pp. 168–173.
Cortez, P., Cerdeira, A., Almeida, F., Matos, T., Reis, J., 2009. Modeling wine preferences by data mining from physicochemical properties. Decis. Support. Syst. 47 (4), 547–553.
D'Astous, A., L'evesque, M., 2003. A scale for measuring store personality. Psychol. Mark. 20 (5), 455–469.
Dinnie, K., 2008. Nation branding. In: Concepts, Issues, Practice. Elsevier, Oxford.
Donkers, B., Verhoef, P.C., de Jong, M.G., 2007. Modeling CLV: a test of competing models in the insurance industry. Quant. Mark. Econ. 5 (1), 163–190.
Forgy, E., 1965. Cluster analysis of multivariate data: efficiency versus interpretability of classification. Biometrics 21 (3), 768–769.
Foster, M., West, B., Francescucci, A., 2011. Exploring social media user segmentation

and online brand profiles. J. Brand Manag. 19 (1), 4–17.
Freling, T.H., Forbes, L.P., 2005. An empirical analysis of the brand personality effect. J. Prod. Brand. Manag. 14 (7), 404–413.
Gelbrich, K., Nakhaeizadeh, R., 2000. Value miner: a data mining environment for the calculation of the customer lifetime value with application to the automotive industry. In: In ECML 2000, LNAI 1810, pp. 154–161.
Geuens, M., Weijters, B., Wulf, K.D., 2009. A new measure of brand personality. Int. J. Res. Mark. 26, 97–107.
Ghosh, S., Dubey, S.K., 2013. Comparative analysis of K-means and fuzzy C-means algorithms. Int. J. Adv. Comput. Sci. Appl. 4 (4), 35–39.
Girvan, M., Newman, M., 2002. Community structure in social and biological networks. Natl. Acad. Sci. U. S. A. 99 (12), 8271–8276.
Gronroos, C., 2007. Service management and marketing. Eur. J. Mark. 15 (2), 3–31.
Gupta, A., Su, B.-C., Walter, Z., 2004. An empirical study of consumer switching from traditional to electronic channels: a purchase-decision process perspective. J. Electron. Commer. 8 (3), 131–161.
Gupta, S., Hanssens, D., Hardie, B., Kahn, W., Kumar, W., Lin, N., Ravishanker, N., Sriram, S., 2006. Modeling customer lifetime value. J. Serv. Res. 9 (2), 139.
Haenlein, M., Kaplan, A.M., Schoder, D., 2006. Valuing the real option of abandoning unprofitable custorners when calculating customer lifetime value. J. Mark. 70, 5–20.
Hamzaoui-Essoussi, L., Merunka, D., Bartikowski, B., 2011. Brand origin and country of manufacture influences on brand equity and the moderating role of brand typicality. J. Bus. Res. 64, 973–978.
Han, C.M., 1989. Country image: halo or summary construct. J. Mark. Res. 26, 222–229.
He, W., Zha, S., Li, L., 2013. Social media competitive analysis and text mining: a case study in the pizza industry. Int. J. Inf. Manag. 33 (3), 464–472.
Heidemann, J., Klier, M., Probst, F., 2012. Online social networks: a survey of aglobal phenomenon. Comput. Netw. 56 (18), 3866–3878.
Hellgeson, J.G., Suphellen, M., 2004. A conceptual and measurement comparison of self-congruity and brand personality. Int. J. Mark. Res. 46 (2), 205–223.
Huang, H.H., Mitchell, V., Rosenaum-Elliott, R., 2012. Are consumer and brand personalities the same. Psychol. Mark. 29 (5), 334–349.
Jain, D., Singh, S.S., 2002. Customer lifetime value research in marketing: a review and future directions. J. Interact. Mark. 16 (2), 34.
Jian, W.U., Guoqun, F.U., 2007. The effects of brand origin country and made-in country on consumers' product evaluations and purchase intention. Front. Bus. Res. China 1 (3), 333–350.
Jipkate, B., Gohokar, V., 2012. A comparative analysis of fuzzy C-means clustering and K means clustering algorithms. Int. J. Comput. Eng. 2 (3), 737–739.
Kamada, T., Kawai, S., 1989. An algorithm for drawing general undirected graphs. Inf. Process. Lett. 31 (1), 7–15.
Kaur, M., Kaur, U., 2013. Comparison between K-mean and hierarchical algorithm using query redirection. Int. J. Adv. Res. Comput. Sci. Softw. Eng. 3 (7), 1454–1459.
Keller, K.L., 2001. Building Customer-based Brand Equity: A Blueprint for Creating Strong Brands, Working Paper Series. Marketing Science Institute, Cambridge, MA.
Keller, K., Richy, L., 2006. The importance of corporate brand personality traits to a successful 21st century business. Brand Manag. 14 (1), 74–81.
Kim, H., Yoon, C.H., 2004. Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. Telecommun. Policy 28, 751–765.
Kim, S.Y., Jung, T.S., Suh, E.H., Hwang, H.S., 2006. Customer segmentation and strategy development based on customer lifetime value: a case study. Expert Syst. Appl. 31, 101–107.
Kim, J., Baek, T.H., Martin, J., 2010. Dimensions of news media brand personality. J. Mass Commun. Q. 87 (1), 117–134.
Kim, C., Jun, J., Lee, H., 2014. Improved churn prediction in telecommunication industry by analyzing a large network. Expert Syst. Appl. 41 (5), 6575–6584.
Krohmer, H., Malär, L., Nyffenegger, B., 2007. The fit between brand personality and consumer's self: the importance of self-congruence for brand performance. In: Paper Presented at the AMA Winter Educators' Conference Brand Identity Communications Track, San Diego, CA.
Kumar, V., Rajan, B., 2009. Profitable management: measuring and maximizing customer lifetime value. Manag. Account. Q. 10 (3), 1–18.
Labbi, A., Berrospi, C., 2007. Optimizing marketing planning and budgeting using Markov decision processes: an airline case study. IBM J. Res. Dev. 51 (3/4), 421–431.
Lee, H.S., Cho, C.H., 2009. The matching effect of brand and sporting event personality: sponsorship implications. J. Sport Manag. 23 (1), 41–64.
Lin, L.Y., 2010. The relationship of consumer personality trait, brand personality and brand loyalty: an empirical study of toys and video games buyers. J. Prod. Brand. Manag. 19 (1), 4–17.
Lycett, M., 2013. Datafication: making sense of (big) data in a complex world. Eur. J. Inf. Syst. 22 (4), 381–386.
Malcolm, M., Dunbar, I., 2012. Market Segmentation: How to Do It and How to Profit from It, 4th ed. John Wiley & Sons, Chichester, UK.
Malthouse, E.C., Blattberg, R.C., 2005. Can we predict customer lifetime value. J. Interact. Mark. 19 (1), 2–16.
Mithas, S., Lee, M.R., Earley, S., Murugesan, S., 2013. Leveraging big data and business analytics. IT Prof. 15 (6), 18–20.
Moilanen, T., Rainisto, S., 2008. How to Brand Nations, Cities and Destinations: A Planning Book for Place Branding. Palgrave Macmillan, London.
Mulhern, F.J., 1999. Customer profitability analysis: measurement, concentration, and research directions. J. Interact. Mark. 13 (1), 25–40.
Mulyanegara, R.C., Tsarenko, Y., Anderson, A., 2009. The Big Five and brand personality: investigating the impact of consumer personality on preferences. J. Brand Manag. 16 (4), 234–247.
Ngai, E.W.T., Xiu, L., Chau, D.C.K., 2009. Application of data mining techniques in customer relationship management: a literature review and classification. Expert Syst. Appl. 36, 2592–2602.
O'Loughlin, D., Szmigin, I., 2006. Customer relationship typologies and the nature of loyalty in retail financial services. J. Mark. Manag. 22 (2), 267–293.
Pappu, R., Quester, P.G., Cooksey, R.W., 2006. Consumer-based brand equity and country-of-origin relationships. Eur. J. Mark. 40 (5/6), 696–717.
Parker, B.T., 2009. A comparison of brand personality and brand user-imagery congruence. J. Consum. Mark. 26 (3), 175–184.
Peppers, D., Rogers, M., 1999. One to One Manager: An Executive's Guide to Custom Relationship Management. Random House, New York, NY.
Phau, I.Y., Lau, K., 2000. Conceptualising brand personality: a review and research proposition. J. Target. Meas. Anal. Mark. 9 (1), 52–69.
Plummer, J.T., 1985. How personality makes a difference. J. Advert. Res. 24 (6), 27–31.
Reinartz, W., Kumar, V., 2003. The impact of customer relationship characteristics on profitable lifetime duration. J. Mark. 67 (1), 77–99.
Rijmenam, M.V., 2014. How Geospatial Data Can Enrich Your Customer Experience and Drive Revenue. (Big Data Start-up (retrieved May 28, 2016 from http://www.bigdata-startups.com/geospatial-data-enrich-customer-experience-drive-revenue/)).
Rohm, A., Swaminathan, V., 2004. A typology of online shoppers based on shopping motivations. J. Bus. Res. 57 (7), 748–757.
Romaniuk, J., 2008. Comparing methods of measuring brand personality traits. J. Mark. Theory Pract. 16 (2), 153–161.
Ryals, L.J., Knox, S., 2005. Measuring risk-adjusted customer lifetime value and its impact on relationship marketing strategies and shareholder value. Eur. J. Mark. 39 (5/6), 456.
Sanyal, S.N., Datta, S.K., 2011. The effect of country of origin on brand equity: an empirical study on generic drugs. J. Prod. Brand. Manag. 20 (2), 130–140.
Shih, Y., Liu, D.R., 2008. Product recommendation approaches: collaborative filtering via customer lifetime value and customer demands. Expert Syst. Appl. 35, 350–360.
Smith, A.C., Graetz, B.R., Westerbeek, H.M., 2006. Brand personality in a membership based organisation. Int. J. Nonprofit Volunt. Sect. Mark. 11 (3), 251–266.
Suganya, R., Shanthi, R., 2012. Fuzzy C-means algorithm — a review. Int. J. Sci. Res. Publ. 2 (11), 1–3.
Sung, Y., Tinkham, S.F., 2005. Brand personality structures in the United States and Korea: common and culture-specific factors. J. Consum. Psychol. 15 (4), 334–350.
Sweeney, J.C., Brandon, C., 2006. Brand personality: exploring the potential to move from factor analytical to circumplex models. Psychol. Mark. 23 (8), 639–663.
Thakor, M.V., Kohli, C.S., 1996. Brand origin: conceptualization and review. J. Consum. Mark. 13 (3), 27–42.
Thomson, M., MacInnis, D.J., Park, C.W., 2005. The ties that bind: measuring the strength of consumers' attachments to brands. J. Consum. Psychol. 15 (1), 77–91.
Turney, P.D., Littman, M.L., 2003. Measuring praise and criticism: inference of semantic orientation from association. ACM Trans. Inf. Syst. 21 (4), 315–346.
Verbeke, W., Dejaeger, K., Martens, D.J., Baesens, B., 2011. New insights into churn prediction in the telecommunication sector: a profit driven data mining approach. Eur. J. Oper. Res. 218, 211–229.
Wedel, M., Kamakura, W., 2002. Introduction to the special issue on market segmentation. Int. J. Res. Mark. 19 (3), 181–183.
Wiertz, C., DeRuyter, K., 2007. Beyond the call of duty: why consumers contribute to firm-hosted commercial online communities. Organ. Stud. 28 (3), 347–376.
Williams, G., 2009. Rattle: a data mining GUI for R. R J. 1 (2), 45–55.
Yasin, N.M., Noor, M.N., Mohamad, O., 2007. Does image of country-of-origin matter to brand equity. J. Prod. Brand. Manag. 16 (1), 38–48.
Zeithaml, V.A., Rust, R.T., Lemon, K.N., 2001. The customer pyramid: creating and serving profitable customers. Calif. Manag. Rev. 43 (4), 118–142.

Lan-lung (Luke) Chiang, got his Ph.D. in Marketing and Management for Hospitality and Services Business, at Iowa State University in 2013. Since 2013 he is an Assistant Professor in the area of marketing at the College of Management, Yuan Ze University, Taiwan. His research fields are consumer behavior, services marketing, destination marketing, and data-mining.