**Using Big Data from Customer Relationship Management Information Systems to Determine the Client Profile in the Hotel Sector**

## Abstract

Client knowledge remains a key strategic point in hospitality management. However, the role that can be played by large amounts of available information in the Customer Relationship Management (CRM) systems, when addressed by the use of emerging Big Data techniques for efficient client profiling, is still in its early stages. In this work, we addressed the client profiling of the data in a CRM system from an international hotel chain, by using Big Data technology and Bootstrap resampling techniques for Proportion Tests. Strong consistency was found on the most representative feature of repeaters being *traveling without children*. Profiles were more similar for British and German clients, and their main differences with Spanish clients were in the *stay duration* and in *age*. For a vacational chain, these results suggest further analysis on the target orientation towards new market segments. Big Data technologies can be extremely useful for analyzing indoor data available in CRM information systems from hospitality industry.

**Keywords**

Big Data, Hospitality Industry, Customer Relationship Management, Client Profiling, Bootstrap Resampling, Hotel Chains.

**1. Introduction**

Customer knowledge is vital for the hospitality industry, and it plays a crucial role in improving the offer with better quality services (i.e., more adapted and customized), the relationship with customers, and the approach of marketing strategies (Adomavicius & Tuzhilin, 2001; Min, Min & Emam, 2002). All of them result in a better customer satisfaction that increases the loyalty and ensures repeating customers, as well as in larger profitability (Tseng & Wu, 2014). Over the last years, this information has been mainly managed in many hotels by proactively gathering and recording customer preferences into the so-called Customer Relationship Management (CRM) systems (Sarmaniotis et al., 2013). CRM have become a key strategy for improving the customer satisfaction and retention, especially in hotels (Padilla-Melendez & Garrido-Moreno, 2014), and they are remarkably beneficial to those organizations by generating large amounts of valuable information about their customers (Kotler, 2002; Nguyen, Sherif & Newby, 2007; Chadha, 2015).

Nevertheless, it has been recently pointed out (Dursun & Caber, 2016) that even advanced analysis techniques, such as data mining, are not yet being adequately used in the hotel industry for the purpose of effectively profiling the customers by using the comprehensive data that are routinely collected with hotel CRM systems. A large amount of information is available nowadays in hotel companies, either internal and structured (from the Property Management and the CRM systems), or external and unstructured (such as opinion platforms, social networks, or geolocalization, among many others). This brings the need to consider powerful tools available from Big Data technologies, which have already been successfully used in other fields such as bioinformatics, healthcare, or finance (George, Haas & Pentland, 2014), to name just a few.

Big Data technologies are providing unprecedented opportunities for statistical inference on massive analysis, but they also bring new challenges to be addressed, especially when compared to the analysis of carefully collected smaller data sets. In Sivarajah et al. (2017), a systematic and illustrative review is presented on the state-of-art analysis of the literature on Big Data techniques and Big Data Analytics, which highlights the key challenges in terms of different data types, data processing, and data management. As pointed therein, descriptive statistics are the simplest form of Big Data analytic methods, and they involve the summarization and description of knowledge and patterns by using simple statistical tests, such as mean, median, mode, variance, or proportions. When scrutinizing the usefulness of Big Data technologies in a new application field, it is necessary to establish well the behavior and scope of basic statistics, before going into more sophisticated analytics such as data mining or advanced machine learning.

In the present work, our main practical objective was to determine the client profile in an international hotel chain by exploiting the overall information in its CRM system. For this purpose, we identified the relevant variable groups, and we analyzed their practical meaning by using Big Data analytics on proportion tests from ratios between repeaters and first-timers. The use of robust and reliable proportion tests in this scenario has been tackled by using Bootstrap resampling techniques, which provides us with clear cut-off tests for decision making even in massive analysis conditions. It was possible for us to obtain two types of implications, namely, those related with the application of Big Data techniques to CRM exploitation, and those related with the results derived from the specific application to this hotel chain.

The scheme of the paper is as follows. In the next section, we introduce the relevance of CRM systems and their applications in the hotel sector, the basics on Big Data techniques and their scope in current analytics for hotel clients, and some relevant studies dealing with client profiling in terms of their repeating behavior. Then, we present the theoretical foundations of the methods to be used in our client profiling study, consisting of proportion tests and on bootstrap resampling. We then present the results on a real database from a large-scale hotel chain. Discussion is established on our and others’ results, and finally, concise conclusions are drawn.

## 2. Literature Review

Our main objective in the present work was to determine the repeater client profile by exploiting CRM systems in hotel chains and using Big Data technologies. For this purpose, we start by presenting a review of the state of art focused on the three main topics developed here, and the CRM system concept is first scrutinized. Recent applications on Big Data technologies are then summarized, both of them in the hotel industry, and finally, we present recent studies analyzing the repeaters versus first-timers profiles.

### 2.1. CRM in the Hospitality Industry

In the last years, CRM has grown in relevance both in the operational and in the strategic points of view. Two of the major reasons for this are the increasing market competitiveness and the lower cost for client retention than for new client recruiting (Petrick, 2004; Yoo & Bai, 2013). Hence, CRM has become a key strategy for personalizing the client experience and for increasing their satisfaction.

The present work deals with CRM systems. A CRM system is a “firm tool that is technology-based for developing and leveraging consumer knowledge to nurture, maintain, and strengthen profitable relationships with consumers” (Elfving & Lemoine, 2012). According to Buttle (2004), a CRM system is a crucial part of a global CRM strategy. Soltani & Navimipour (2016) stated that CRM systems provide the infrastructure that facilitates the construction of long-term relationships with customers. Some examples of the functionality of CRM systems are sales force automation, data warehousing, data mining, decision support, and reporting tools (Katz, 2002; Hendricks, Singhal & Startman, 2007; Soltani & Navimipour, 2016).

For the hospitality sector, several studies consider CRM as one of the best strategies for improving the company results and for ensuring long-term survival (Sigala, 2005; Abu-Kasim & Minai, 2009; Wu & Li, 2011; Keramati, Mehrabi, & Mojir, 2010; Kim & Choi, 2010). Accordingly, CRM systems are nowadays a fundamental tool in the hotel sector, especially when properly implemented, due to the large amount of data that hotels integrate from their clients. These data could be turned into useful knowledge (Dev & Olsen, 2000; Kotler, 2002; Lin & Su, 2003; Nguyen et al., 2007; Nasution & Mavondo, 2008; Chadha, 2015), and the implementation of CRM systems allows us to identify the host behavioral patterns and to retain them in the long term (Papastathopoulou et al., 2007; Verdugo, Oviedo-Garcia & Roldan, 2009; Chadha, 2015).

It is evident that customer loyalty and profitability are correlated (Payne & Frow, 2005). Therefore, one of the main assumptions of CRM systems is that satisfying and creating long-term relationships with profitable customers enhances the business success of the company (Wu & Lu, 2012). However, the role that large amounts of information currently available in CRM systems can play on efficient client profiling has not been studied enough yet, even for simple and well known statistical descriptions. In addition, there is evidence that advanced analysis techniques are not yet being properly used in the hotel industry to effectively profile customers from comprehensive data collected via hotel CRM systems (Dursun & Caber, 2016). Hotels are not fully exploiting the potential of CRM systems, but there is strong interest and ongoing work towards their successful implementation (Padilla-Melendez & Garrido-Moreno, 2014). This way, both the CRM-effort efficiency and the company competitiveness could be dramatically increased.

### 2.2. Big Data in the Hospitality Industry

Big Data is drastically changing the hotel sector management and the client-to-business relationship, by making easier the decision-making process from large amounts of data (Fox & Do, 2013). Nowadays, the technological bases of both the tourism organizations and the hoteliers make relevant that marketers and managers improve their access to data intelligence to make its best use (Peter, 2014). These professionals have heavily invested in recent years in organizing strong scientific teams and including statisticians and database experts who are well equipped to build and analyse the contents of their Data Warehouses (Ramos, Martins, Serra, Lam, Cardoso, Correia & Rodrigues, 2017). Though human analysis is often required, today Big Data can enhance the decision making and increase the organizational output from five possible approaches, namely, descriptive analytics, inquisitive analytics, predictive analytics, prescriptive analytics, and preemptive analytics. Most Big Data analytics are descriptive and exploratory in nature, but even simple descriptive statistics allow businesses to discover simple and clear patterns that become extremely useful for decisions.

The hospitality industry has become an information intense sector, where large data volumes have been stored with practical applications which are not so widespread. With the arrival of Big Data, it is possible to manage such data in order to achieve the objectives and to transform the information into knowledge (Xian, Schwartz, Gerdes & Uysal, 2015). Data are stored in very different formats, and their analysis becomes a complex task due to their heterogeneity, going from structured data in conventional databases (from Property Management Systems and CRM systems) to semi-structured and unstructured data. Furthermore, the available information systems often can include meta-search generated data, e.g., Tripadvisor, Kayak, Trivago, or social networks such as Facebook, Twitter, or LinkedIn (Santana-Cerdeña, Ramos, Bobur, 2014; Ramos et al., 2017).

The hotel industry is starting to use Big Data technologies mainly in product sales, social media and online behavior of customers, as well as offline data retrieval and analysis (Zhang, Shu & Wang, 2015). Some examples are tourist´s location (Silva & Mateus, 2003; Hjorth, 2012; Vu, Li, Law & Ye., 2015), blogs (Litvin, Goldsmith, & Pan, 2008; Tseng, Wu, Morrison, Zhang, & Chen, 2015), photography (Balomenou & Garrod, 2014), internet behavior (Rong, Vu, Law, & Li, ,2012), search engines (Pan & Li, 2011), and Online Travel Agencies (Ramos et al., 2017), to cite just a few. There is a growing interest in the hospitality field to exploit user-generated data and gain insight into research problems that have not yet been well understood using conventional methods (Ye, Law & Gu, 2009; Yang, Pan & Song, 2013). The most used Big Data method in this setting is Text Analytics for information retrieval, and it usually involves machine learning, statistical analysis, and computational linguistics (Özköse, Arı, & Gencer, 2015). The main functionalities of Big Data and CRM systems in hotel management focus on Revenue Management and on Marketing, and their use includes forecasting, pricing, and benchmarking (Pan & Yang, 2016; Haynes, 2016; Noone 2016; Song & Liu 2017; Ramos et al., 2017).

However, the growing interest in Big Data for dealing with external and unstructured data from clients has brought attention to the fact that large hotel chains can also access a huge amount of internal and structured data through their CRM systems. These systems may not often be deeply exploited for client knowledge purposes, even though simple statistical analysis in Big Data can yield significant and powerful pattern descriptions. Whereas some other studies have paid attention to Big Data approaches with advanced analytics and to its usefulness on internal data generated by hotel clients (Lee, Hwang, Jo & Kim, 2016), the present study aims to gain insight specifically on the client profiling and restricting ourselves to simple Big Data analytics tools, such as proportion tests.

### 2.3. Client Profiling in First-timers vs Repeaters

Researchers have pointed out the relevance of understanding the differences between first-timers and repeaters in the hospitality and tourism industry (Opperman, 1997; Anwar & Sohail, 2004; Petrick, 2005; Morais & Lin, 2010). Marketing researchers suggested that understanding the differences between first-timers and repeaters can provide us with an excellent basis for market segmentation (Formica & Uysal, 1998; Ryu & Han, 2011). Several studies have identified both tourist profiles, mostly focused on touristic destinations (Opperman, 1997; Fallon & Schofield, 2003; Lau & McKercher, 2004; Wang, 2004; Kim & Prideaux, 2005; Li, Cheng, Kim, Petrick, 2008; Morais, 2010; Vu, Le, Law & Ye, 2015; Tasci, 2016), festival and cultural events (Formica & Uysal, 1998; Anwar & Sohail, 2004), restaurants (Ryu & Heesup, 2011), cruises (Petrick, 2004), or whitewater rafting (Fluker & Turner, 2000). These differences are fundamental to develop effective business and marketing strategies, as well as to understand client motivation and to build theoretical knowledge on decision-making (Oppermann, 1997; Lau & McKercher, 2004; Petrick, 2004). However, few academic references can be found in the hotel sector (Kim, Knutson, & Vogt, 2014).

It has been suggested (Chakravarti & Day, 1991) that different consumer profiles should be determined and identified in terms of different features. Some authors have pointed out the relevance of behavior variables, such as past buying patterns (Wind & Lerner, 1979; Kim et al., 2006; Bayer, 2010), and even more emphasis has been made on the relationship between the repetition habit of the client and the loyalty (Yoo & Bai, 2013). Moreover, repeaters can become efficient communication channels for relatives, friends, colleagues, and other potential consumers (Petrick, 2004).

On the other hand, Tasci (2016) shows that loyal consumers are different from others in terms of socio-demographic, psychographic, and behavioral features. Hence, demographic factors (such as education, gender, and age), as well as travel behavior variables (including purpose of travel), have been found to strongly influence consumer loyalty (Homburg & Giering, 2000; Skogland & Siguaw, 2004), and they pose notable differences between first-timers and repeaters (Oppermann, 1997; Lau & McKercher, 2004; McKercker, 2004; Li et al., 2008). When socio-demographics were analyzed, significant differences were found based on age, spending patterns, length of stay, and nationality (Gitelson & Crompton, 1984; McKercher, 2004; Li et al., 2008).

Our study focuses on the profile of repeaters in an international chain. McKercher (2004) and Lau & McKercher (2004) classified travellers to holiday destinations as either first-timers or repeaters. Many holiday destinations rely heavily on repeated visitations (Fallon & Schofield, 2003; Anwar & Sohail, 2004), hence, the results of this differentiation can be very relevant for this sector.

## 3. Statistical and Data Analytics Methods

### 3.1. Big Database

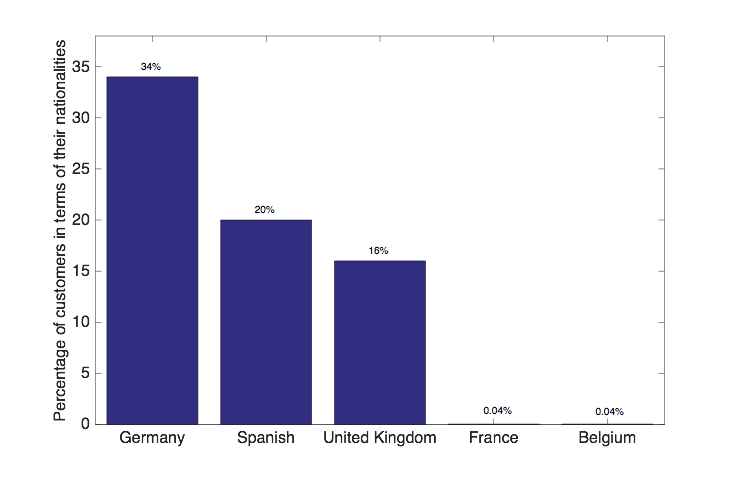
The database used in this study was assembled from a CRM system routinely used by a widely known international hotel company with international scope. The Information Technology unit of the company supervised the access to the information, and a business object was created upon the CRM system universe allowing us to scrutinize the stored features (variables) and to pre-filter the possibly useful ones. A Web Intelligence document was specifically created in the system to support this preliminary analysis.

The information of 4,935,806 different clients (those who stayed overnight at the hotels in the chain) was recorded in the CRM system during years 2013 and 2014. Table 1 describes the variables used in this work. A total of 18 variables as well as their corresponding categories were analyzed, including among them demographic variables, behavioral variables, and transactional data. In addition, the information generated by the hotel chain was also considered as relevant for the present study: guests grouped into repeaters (those clients who stayed more than once at the hotels in the chain) or first-timers and a client clustering used by the chain.

Some studies have indicated that travellers with different nationalities have different motivations for travelling (Mok & Armstrong, 1998; Kim & Lee, 2000; Seddighi, Nuttall & Theocharous, 2001; Kozak, 2002) and that they may show differences in their behavioral characteristics (Kim & Prideaux, 2005). Based on that, we focused on the three nationalities with the largest number of clients in our chain, namely, Germany, Spain, and the United Kingdom (UK), which represented up to 70% of the total clients of this hotel chain (see Figure 1).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Group of variables** | **Variables** | **Categories** | **Type** | **Examples** |
| **Demographic** | Age | 1 | Numerical |  |
| Gender | 3 | Categorical | Female, male, unknown |
| Civil status | 3 | Categorical | Single, with partner, unknow |
| Country of residence | >30 | Categorical | Spain, Germany |
| State | >150 | Categorical | Berlin, Madrid |
| **Behavioral** | Preferred destination | 10 | Categorical |  |
| Motivation | >150 | Categorical | Leisure, business |
| Family holidays | 3 | Categorical | Yes, not, unknown |
| Traveling with a partner | 3 | Categorical | Yes, not, unknown |
| Traveling with children | 3 | Categorical | Yes, not, unknown |
| How did client know about the chain hotel? | >30 | Categorical | Internet, friends, others |
| **Generated by the hotel chain** | Repeaters vs. first timers | 2 | Binary | Yes or not |
| Cluster 1 done by the hotel |  | Categorical |  |
| Cluster 2 done by the hotel |  | Categorical |  |
| Cluster 3 done by the hotel |  | Categorical |  |
| **Transactional Data** | How was the customer registered? | >40 | Categorical | Cardex, call center |
| Entity in the last visit | 8 | Categorical | Agency, B2B |
| Type of price applied to the customer | 3 | Categorical | Promotion |

***Table 1***. *Variable description.*

******

***Figure 1***. *Percentage of customers in terms of several of their nationalities.*

### 3.2. Proportion Tests and Bootstrap Resampling

In this study, we need to compare an overall high number of categories distributed among the features stored in the CRM system. We limited our study to the most usual feature types in the CRM system, i.e., those ones described either by dichotomous or by categorical values, possibly (and very often) in the presence of missing values. We worked with the whole set of categories in terms of two groups which are of our special interest, i.e., repeaters (Group 1) vs first-timers (Group 2).

As an example, think on a simple feature, such as *gender*, for which we have three possible values in the database, namely, female (value = 0), male (value = 2) and unknown (value = 3). In this case, we build a simple vector consisting of three elements, as described following. If we denote by the female proportion in Group 1, and by the female proportion of female in Group 2, we can define the Category Proportion Difference (CPD) as

(1)

Accordingly, if (for example) our analysis yielded , this would indicate that this value of the feature should be more present in Group 1 than in Group 2, and as far as Group 1 is repeater client, it could be read as *women are more prone to be repeater clients*.

We can build the same statistical description for male and for unknown values of the feature, namely, and . Note that these proportions are quite related among them, nevertheless, we are scrutinizing explicitly all the possible values in terms of dichotomic categories. Note also that in the case of unknown or missing data, we should read as *unknown gender is more prone to be present in the database for repeater clients*. Hence, for this example simple feature, we have a three dimensional vector of difference of proportions, which will be denoted as

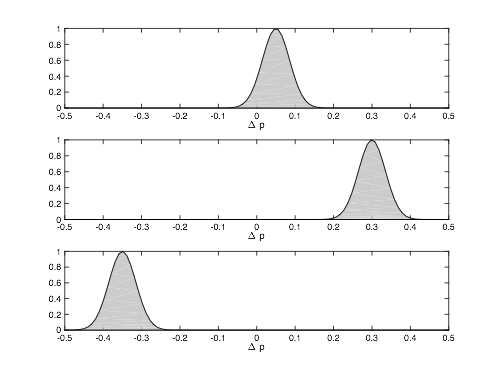
(2)

and it is called here the Feature Proportions Difference Vector (FPDV). The FPDV quantification and visualization provides us with a clear, exhaustive, and simple description of the trends in the values of this feature, and its extension to categorical features is straightforward.

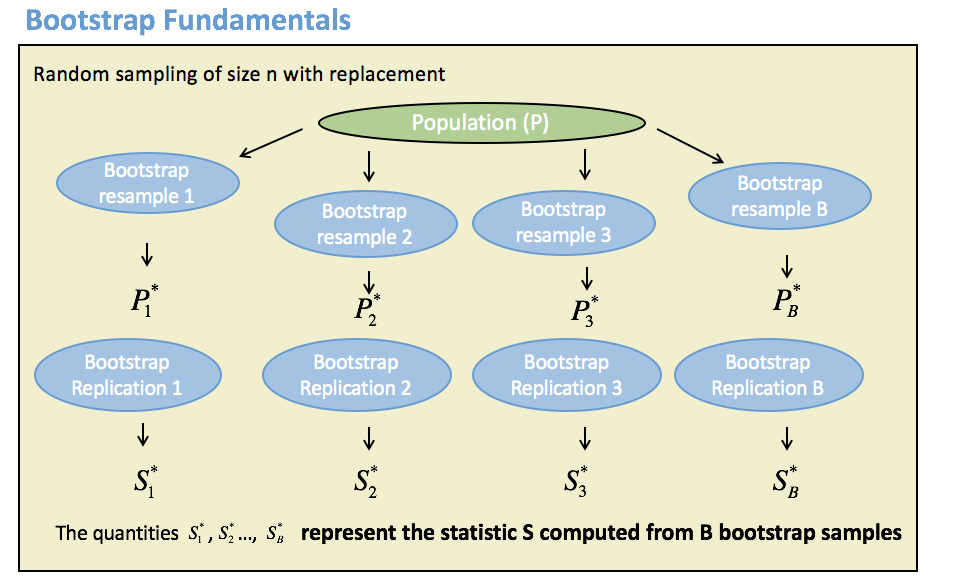
When the proportions of the i-th category is quite similar between groups, their CPD, denoted from now as will be about zero. Nonzero values will deserve special attention, as they are indicative of differences between both groups. At this point, we need to establish a simple-to-use statistical test for determining whether a value of is large enough to indicate a significant difference in this category between the compared groups. Therefore, the statistical test that we need has the following hypothesis:

1. *Null hypothesis*, H0: , and there is no proportion difference between groups for the i-th category.
2. *Alternative hypothesis*, H1: and there is a proportion difference between groups for the i-th category. If (if ), then the proportion of the i-th category is significantly larger in Group 1 (in Group 2).

This hypothesis test is depicted in Figure 2. The described CPD hypothesis test will be intensively used in large databases, as in the case of our CRM features analysis. The proportion test for analyzing could be dealt with simple statistical tests, however, it is well known that proportion tests should be dealt with special caution (Sá, 2003), and more, the definition of parametric tests for can be sensitive in classical statistics, for instance in terms of the actual assumed distributions. For these reasons, we propose here to use a new method for estimating clear cut-off hypothesis tests for CPD by using bootstrap resampling, as described next in short.



***Figure 2***. Hypothesis test for the i-th CPD.



***Figure 3***. Bootstrap resampling fundamentals for providing with a nonparametric hypothesis test for CPD in each category in the database.

The bootstrap resampling is a widely extended and common statistical technique which relies on random sampling with replacement (also known as the plug-in principle) to provide with statistical tests. The plug-in principle says that the empirical distribution function of a statistic can be used as an approximation for the actual distribution function (Efron & Tibshirani, 1994). The rationale of the bootstrap resampling method is that if we want to make some inference of a population in terms of some decision statistic whose calculation is known, but its actual statistical distribution is not easy to obtain, we can resample the population data and make the inference on the resample.

Figure 3 shows the general approach to bootstrap resampling. In our case, we want to provide a nonparametric hypothesis test for the CPD statistic in each category in the database. Given a population, we sample it with replacement a number of times B, which needs to be large enough to represent the tails of the statistical distributions (typically B = 50, 200, or 1000, depending on the application). These statistical copies of the original population are known as bootstrap resamples. The calculation of the decision statistic (in this case, the CPD for a category) is made now on each resample, hence yielding a new value for it, known as *bootstrap replication*. The B replications of the statistics can be used to build a histogram, which stands for an estimation of the actual CPD probability density function, and then, simple methods, such as ordered statistics, can be used to determine the hypothesis test. Note in Figure 3 and in the following that symbol \* is used to identify the quantities that have been estimated throughout the bootstrap process from the original ones, these last often called the *empirical* statistics.

### 3.3. Big Data and Map-Reduce Implementation

The increasing amount of data that is generated, along with the need to extract useful knowledge from them, has emerged into a large number of problems in the last years. These problems are mainly due to the nature of these available data, since the tools used so far were not applicable to these datasets. In particular, these problems affect different areas, such as hardware storage and accessibility of these data, database management, pre-processing, simple analysis, or automatic extraction of hidden patterns.

In order to address all these problems, a plethora of techniques have emerged that allow, in a much more efficient way, to extract value from data that were difficult to address, either by the large size of the available database, or by its low quality, or for being a union of diverse data sources with different nature. The framework encompassing this whole set of tools and solutions tends to be referred to as *Big Data*. In this article, we will use those tools that allow us to pre-process and analyze large volumes of available data, since their size makes it infeasible to analyze them using the classic tools. In particular, one of the tools that can do this is known as *map-reduce* (Jeffrey & Ghemawat, 2008), an algorithmic framework that takes advantage of both parallel hardware architecture and distributed file systems. In our area of ​​knowledge, one of the Big Data challenges aimed to extract useful knowledge from large amounts of data is to create or adapt statistical or machine learning techniques to the parallelizable framework known as map-reduce. This map-reduce framework consists mainly of two basic operations, map and reduce. The map function is distributed and executed in parallel in the different computing nodes that constitute the hardware architecture. The purpose of this map function is to transform the available data format (possibly unstructured) into a structured data form, which is known as *key-value* and that allows us to solve the problem. The reduce function, also executed in parallel, takes as input the outputs of a set of map functions and summarizes all the key-value pairs belonging to the same key in a single key-value pair. Therefore, the obtained result returns as many key-value pairs as keys exist. Because the data has to be transformed into this rigid key-value structure, not all problems seem to be resolved efficiently with this map-reduce framework. In addition, most of the algorithms that allow advanced data analysis (as are those in the machine learning field) are not easily parallelizable (embarrassingly parallel), and they require great effort to be implemented in the map-reduce form.

In this paper, we propose to apply a combination of classical statistical techniques (described in Subsection 3.1) that are embarrassingly parallel in the map-reduce framework, in order to extract useful knowledge from the large available data collection. For its implementation, each bootstrap resampling block is distributed to each available parallel computing node, and the proportion test is applied as a map function. The union or sum of the outputs of each proportion test constitutes the reduce function, thus obtaining the desired advanced data analysis. Before this analysis, a necessary preprocessing is applied through map-reduce procedure in order to prepare the appropriate variables and the necessary information for the study.

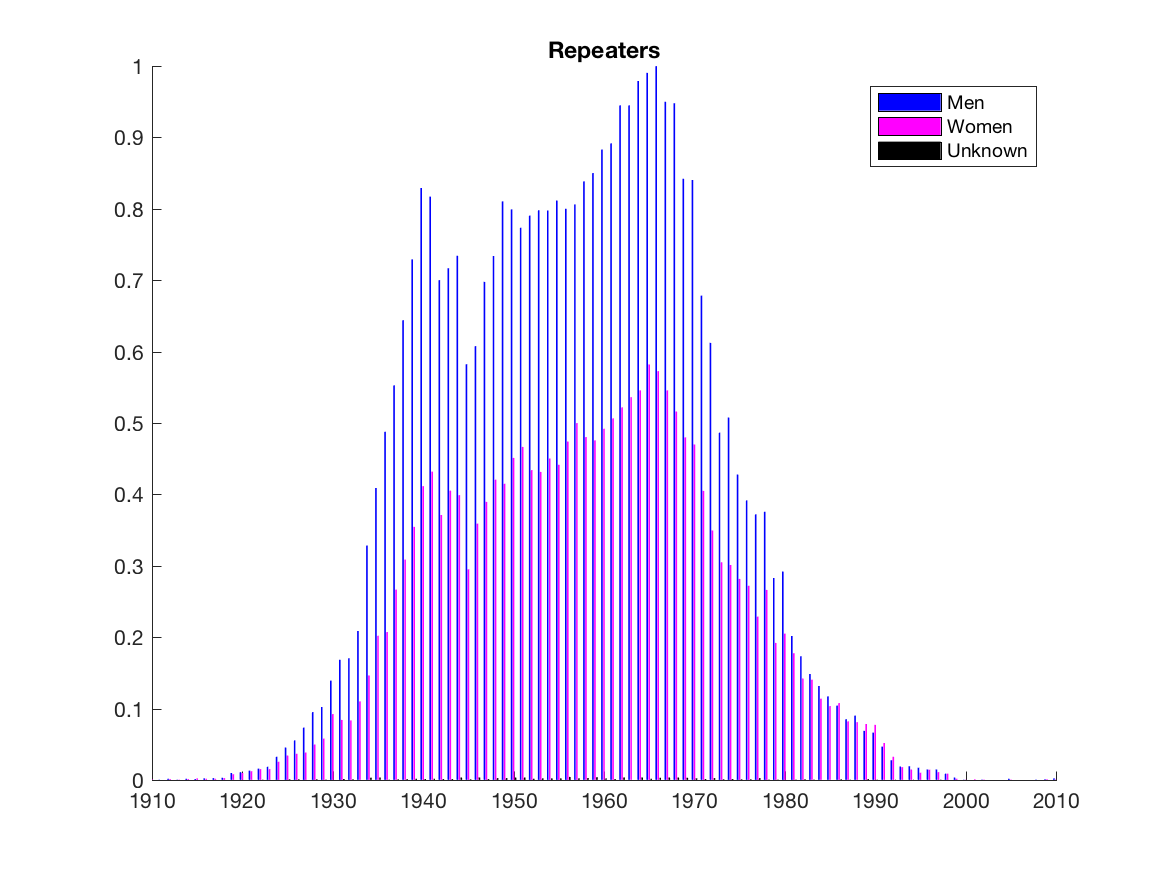
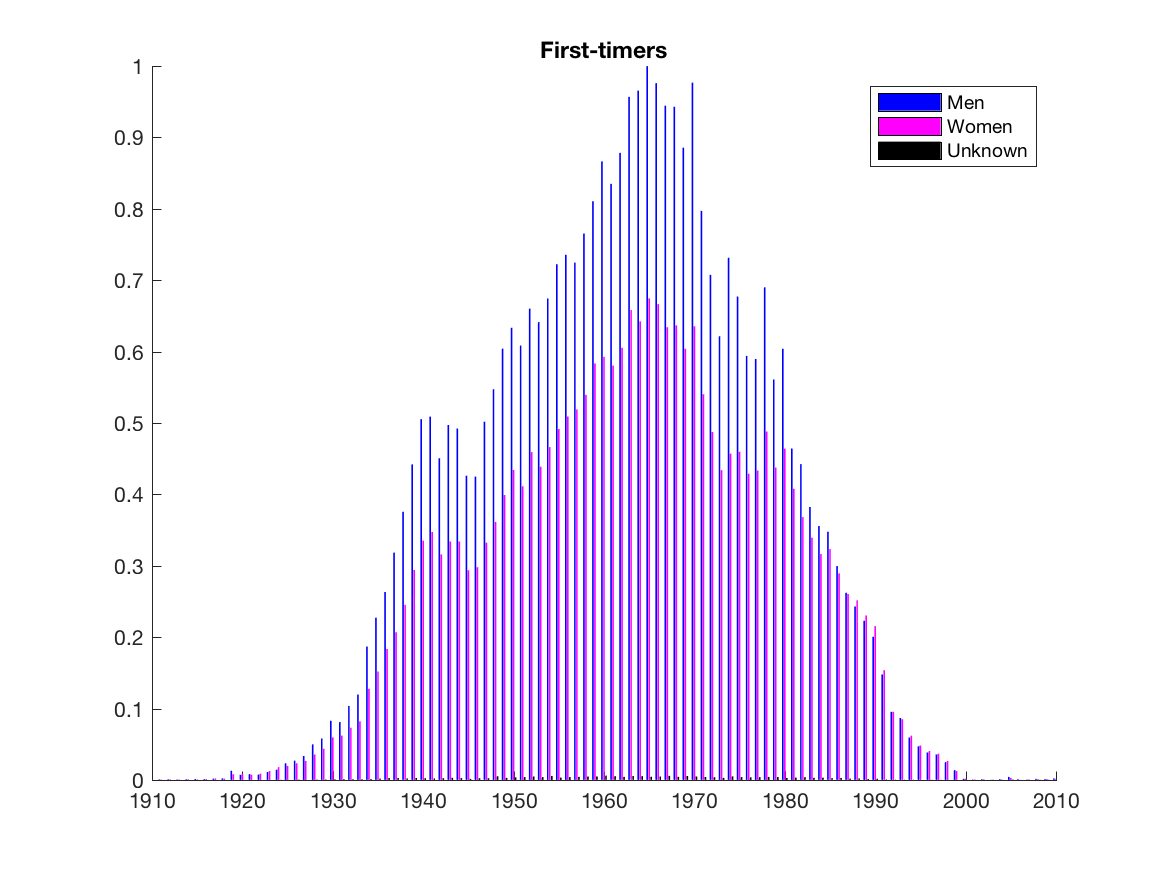
## 4. Results

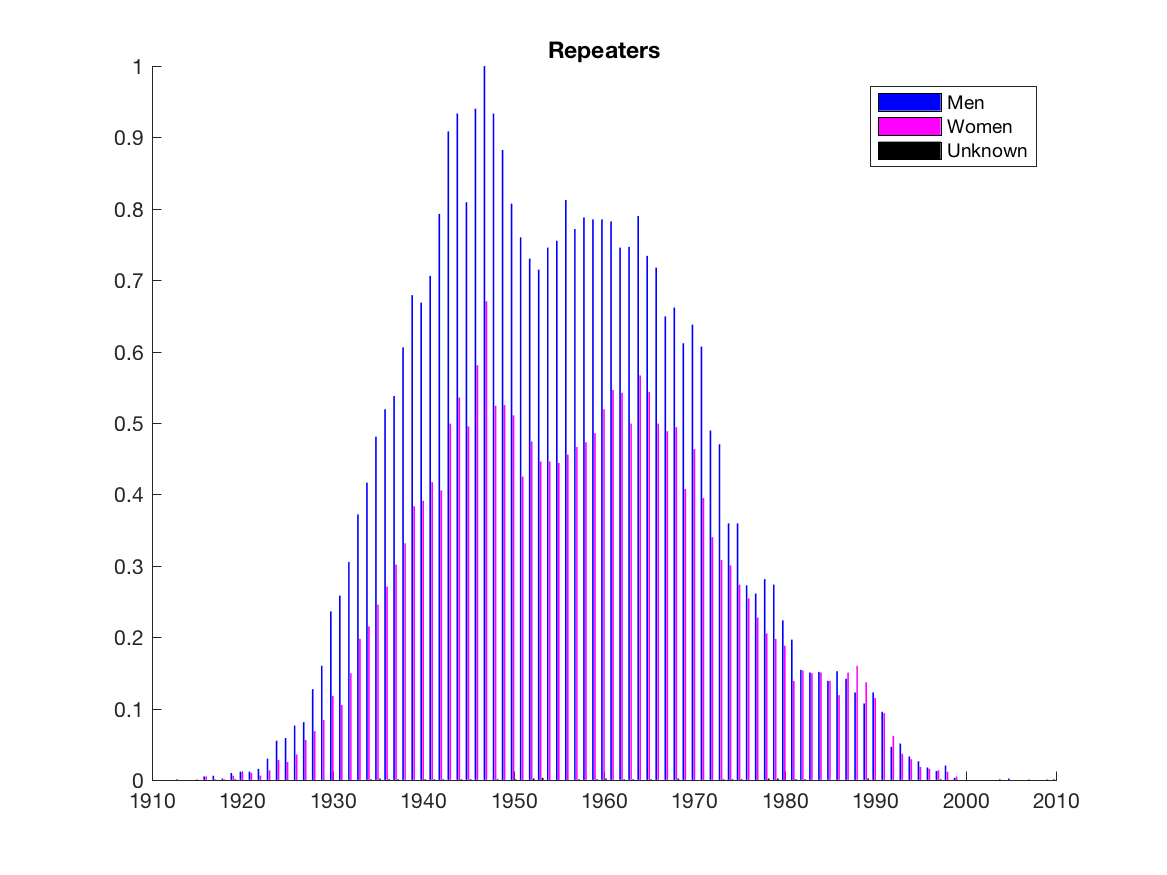
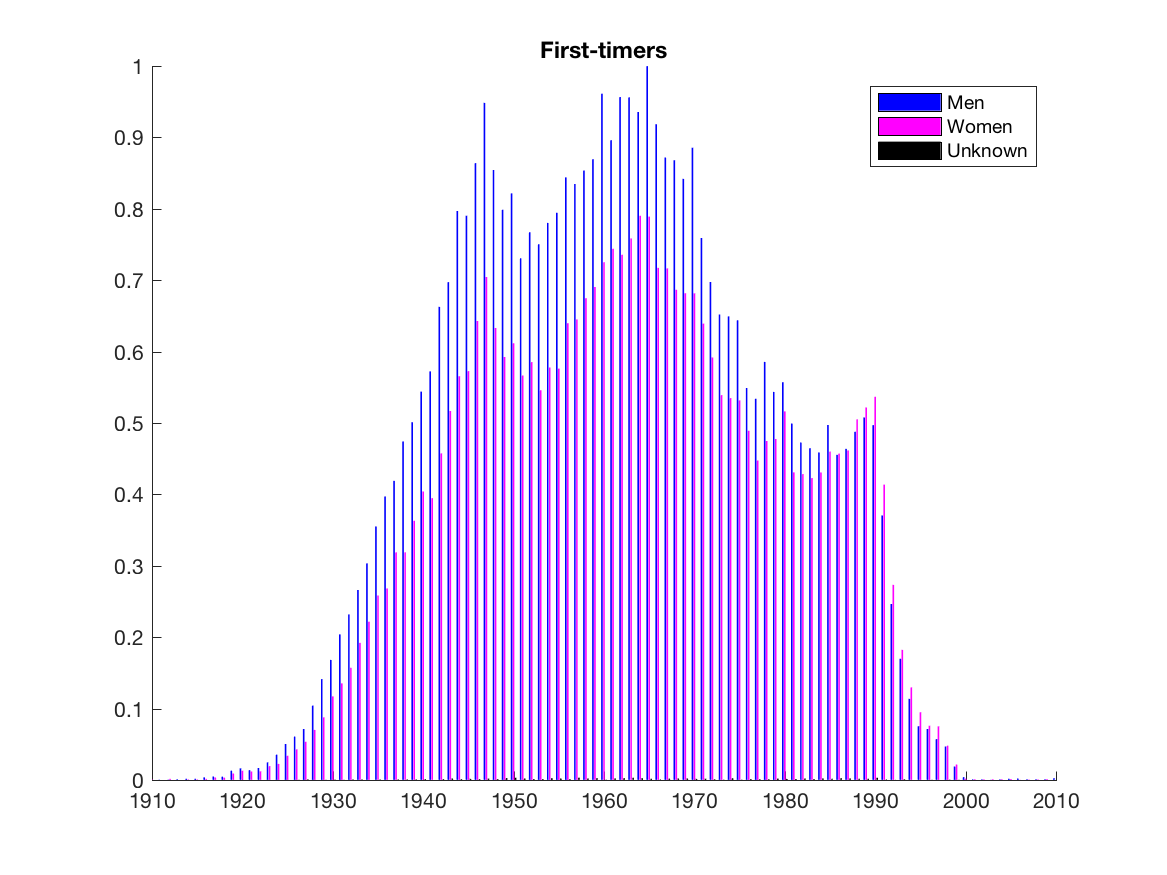
### 4.1. Some Descriptive Statistics

As previously described, data were recorded from 4,935,806 different clients in the CRM system during the years 2013 and 2014, and the goal of this work consisted of modeling the profile of first-timers and repeaters within this population, as it was found usual that many clients stayed in the chain hotel several times. Towards that end, we analyzed first-timers vs repeaters considering their nationality. The whole analysis on the complete dataset took about 15 minutes on a MacBook Pro(R) (3.5 GHz, Intel Core i7), whereas it can be expected to be reduced to just some few seconds when using instead a graphic processing unit (GPU) or another computation server. For each of the selected nationalities (Germany, the UK, and Spain), the age distribution is represented in terms of the gender (men, women, and unknown) for first-timers and repeaters in Fig. 4. Table 2 also summarizes the mean and the standard deviation. In general, first-timer men were older than repeaters, and the age for British citizen was lower than for the other nationalities.

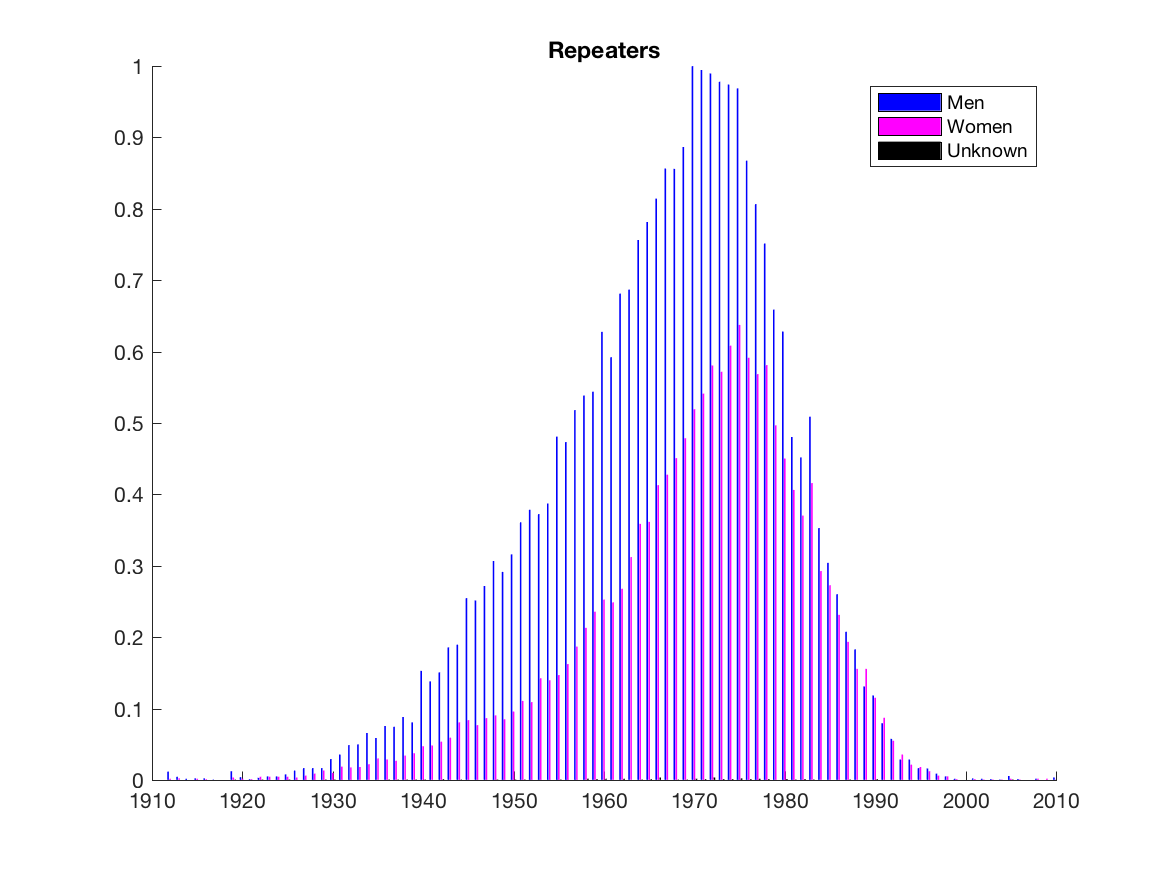
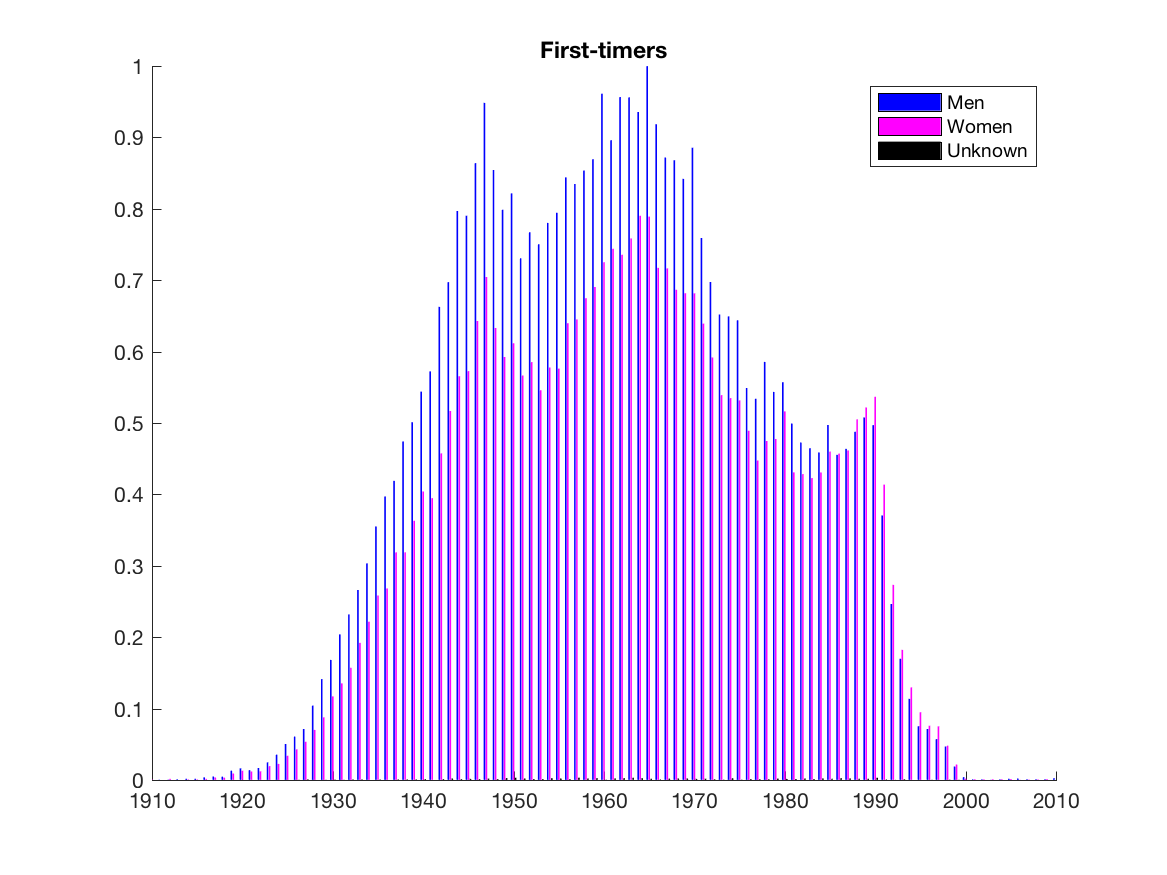
|  |  |  |  |
| --- | --- | --- | --- |
| Gender | Nationality | First-timers | Repeaters |
| Men | German | 1962.18 +- 14.64 | 1956.80 +- 13.86 |
| Spanish | 1966.19 +- 14.46 | 1966.65 +- 12.61 |
| British | 1961.14 +- 16.26 | 1955.17 +- 14.55 |
| Women | German | 1963.19 +- 15.28 | 1958.16 +- 14.48 |
| Spanish | 1968.32 +- 14.86 | 1970.31 +- 12.68 |
| British | 1962.88 +- 16.69 | 1957.39 +- 15.22 |
| Unknown | German | 1960.88 +- 18.16 | 1956.02 +- 15.37 |
| Spanish | 1949.85 +- 24.20 | 1964.12 +- 17.09 |
| British | 1963.92 +- 17.56 | 1957.62 +- 20.33 |

**Table 2**: Mean and standard deviation of birth year in terms of gender, for first-timers and repeaters. First (second, third) row refers to German (Spanish, British) customers.

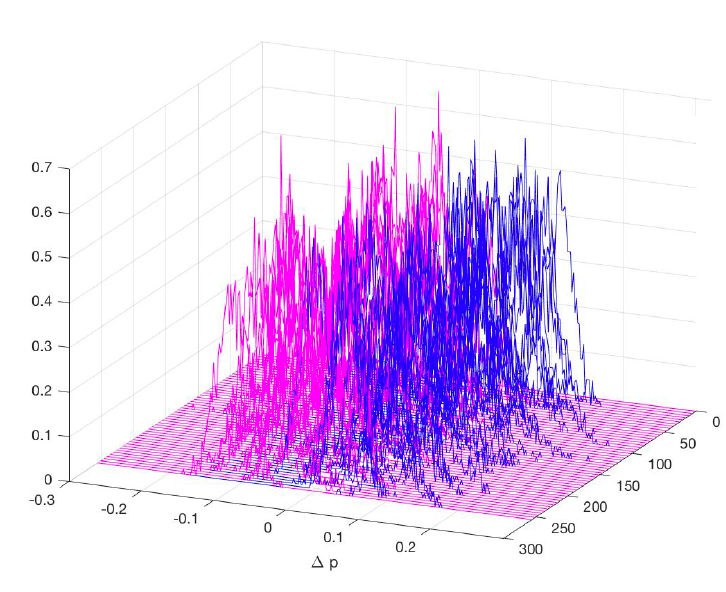
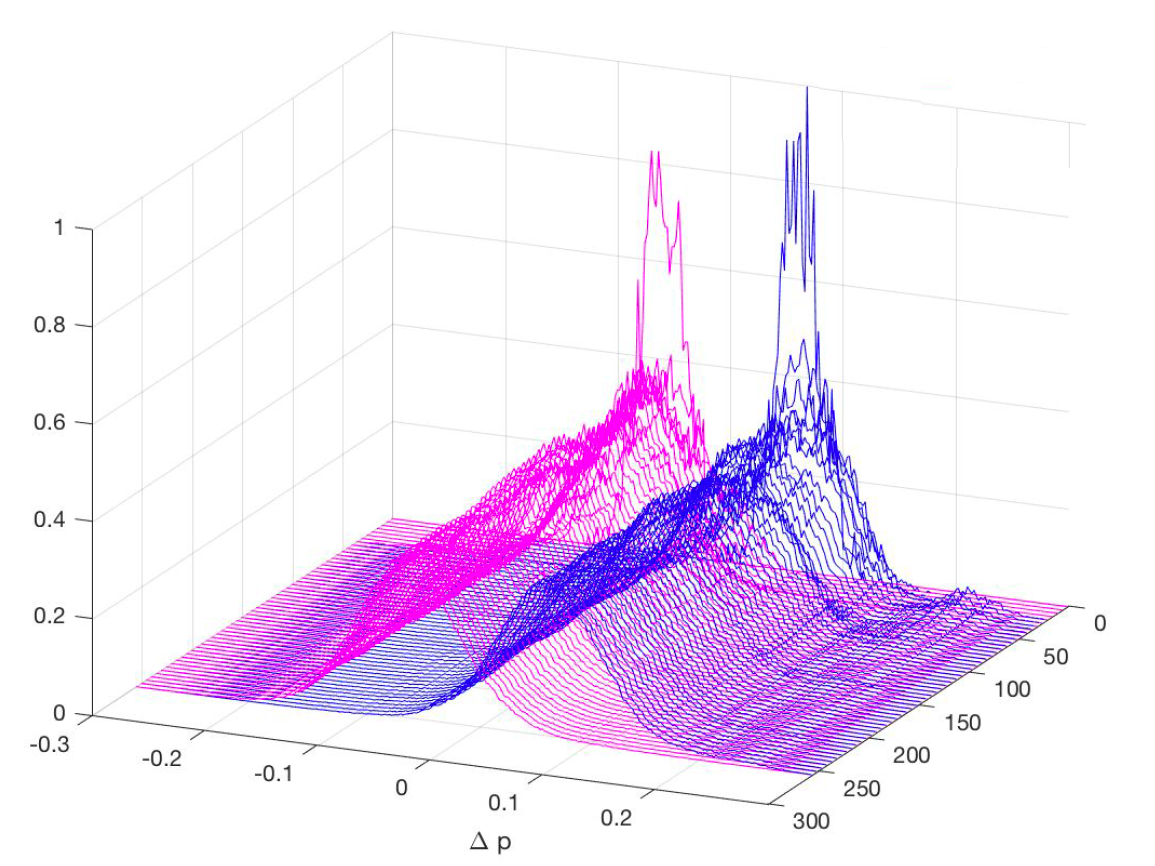
(b)

(c)

***Figure 4***. Birth year distribution in terms of gender (men, women and unknown) for repeater and first-time visitors: (a) Germany; (b) UK; and (c) Spain.

### 4.2. Advantages of Big Data Hypothesis Tests

(a) (b)

***Figure 5***. Histograms yielding the use of hypothesis testing for CPD statistic in an example category (gender in Spain, men in blue, women in pink): (a) Normalized histograms estimated for each chunk in the map-reduce procedure; (b) Aggregated normalized histograms after each chunk in the map-reduce procedure.

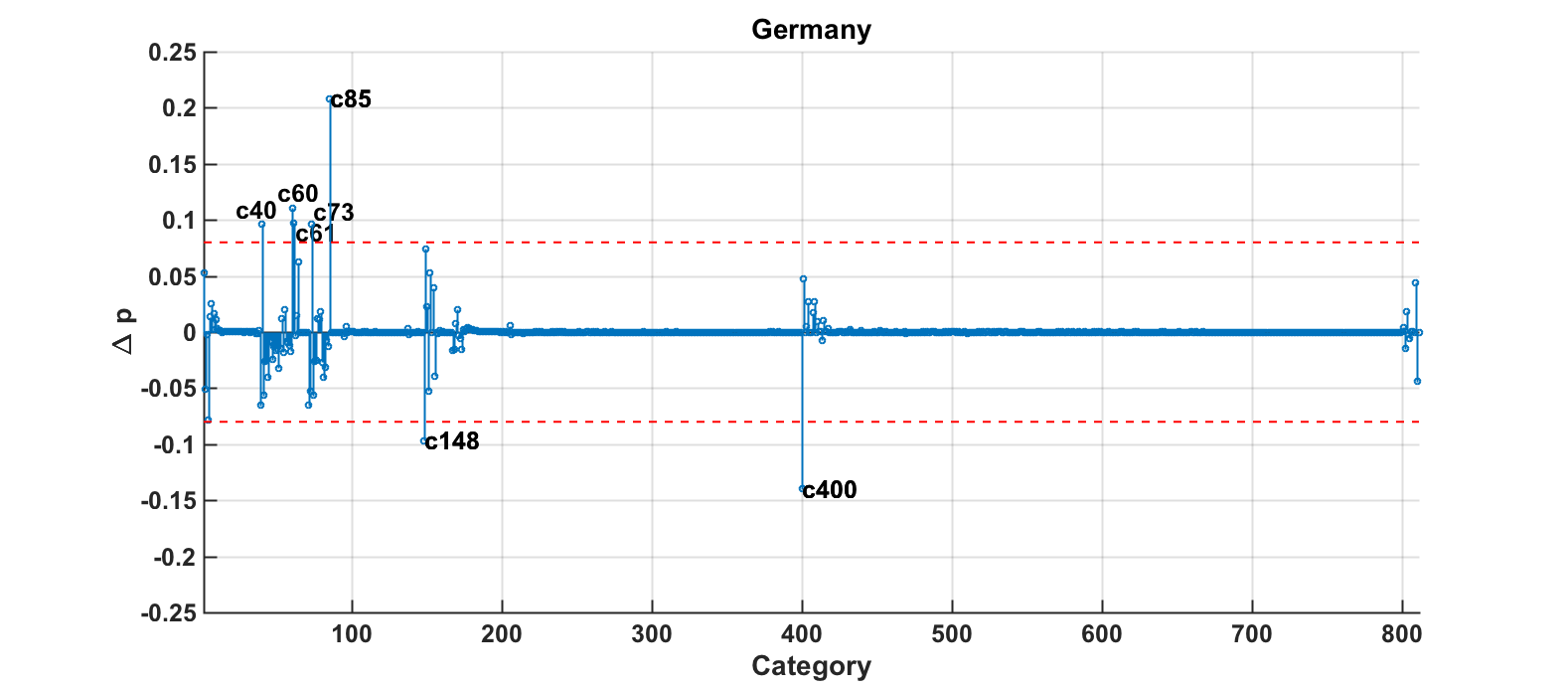
Hypothesis tests based on probabilities are known to be sensitive in terms of statistical robustness. This can be checked in Fig. 5(a), which represents the histograms estimated inside each map-reduce chunk for the CPD statistic of *gender* feature (*unknown* value is not represented in this example). Although the use of these statistical distributions for hypothesis testing separately in each chunk would show a trend to identify these variables as significant, there is still strong variability in these distributions, so that they do not represent a good-quality representation of the probability density function of the CPD statistics. This is surprising even given the large size of each chunk (200,000 cases).

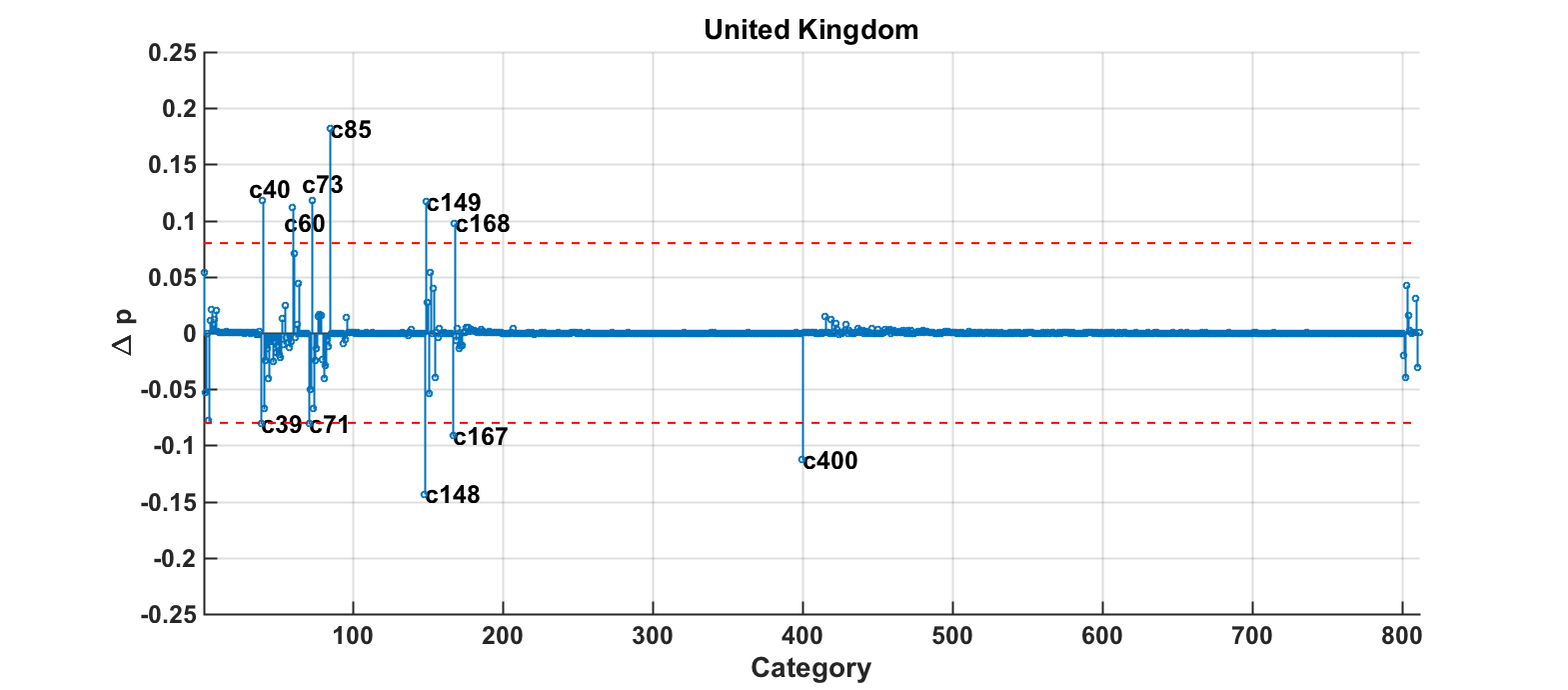
However, Fig. 5(b) shows the aggregated estimation of the probability density functions from the accumulated averaged histograms after each chunk. Whereas the first 10 chunks still provide a quite noisy and fragmented density estimation, it soon becomes a smoother estimation, and it reaches convergence (highly similar estimations) after about half of the chunks have been included in the averaging. This holds even while the distribution estimation is capturing subtle but present multimodalities, which justifies the use of nonparametric statistical methods such as the bootstrap resampling. These observations make evident that even simple statistics are efficiently dealt with Big Data analytics (panel b) than with conventional approaches (panel a).

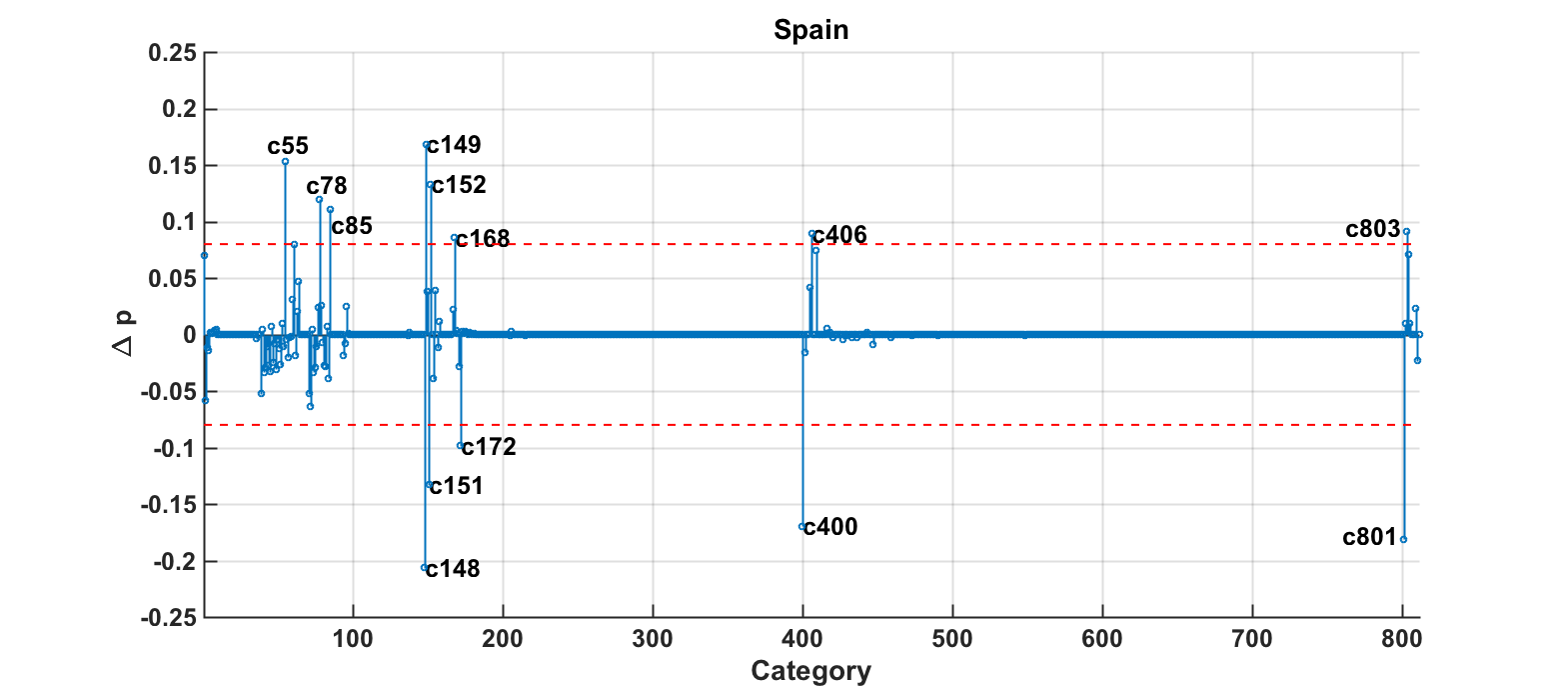
### 4.3. Chromosome and Spiderweb Results for Proportion Tests

We obtained a nonparametric hypothesis test for the CPD statistics in each category in the database and for each nationality. Hypothesis tests for the i-th CPD were calculated by using the bootstrap resampling technique for the proportion tests. Two different visualization techniques were used for the large number of categories scrutinized in the Big Database, namely, the so-called chromosome proportions plot and the spiderweb representation. The first one shows all the proportion rates for all the variables except for motivation due to visualization purposes (more than 150 categories) in three different nationalities (Germany, the UK, and Spain), as seen Fig.6. All the values significantly higher (lower) than 0 indicate that a significant difference exists between the proportions of the i-th category for repeaters (first-timers). Red-dotted lines are used as a threshold to identify the most relevant features with a detailed and extensive description in terms of all the scrutinized categories, so that the higher its absolute value, the more relevant that feature is. We can see that interesting features often are viewed as clusters or interest regions. Alternatively, a joint visualization for defining the clients’ profile based on just the most relevant features are represented by a spiderweb plot (see Fig. 7), which can provide the managers with a coarse and rough view of the profile. Both of these representations are complementary and show different details to the manager for analyzing the CRM features from a business point of view.

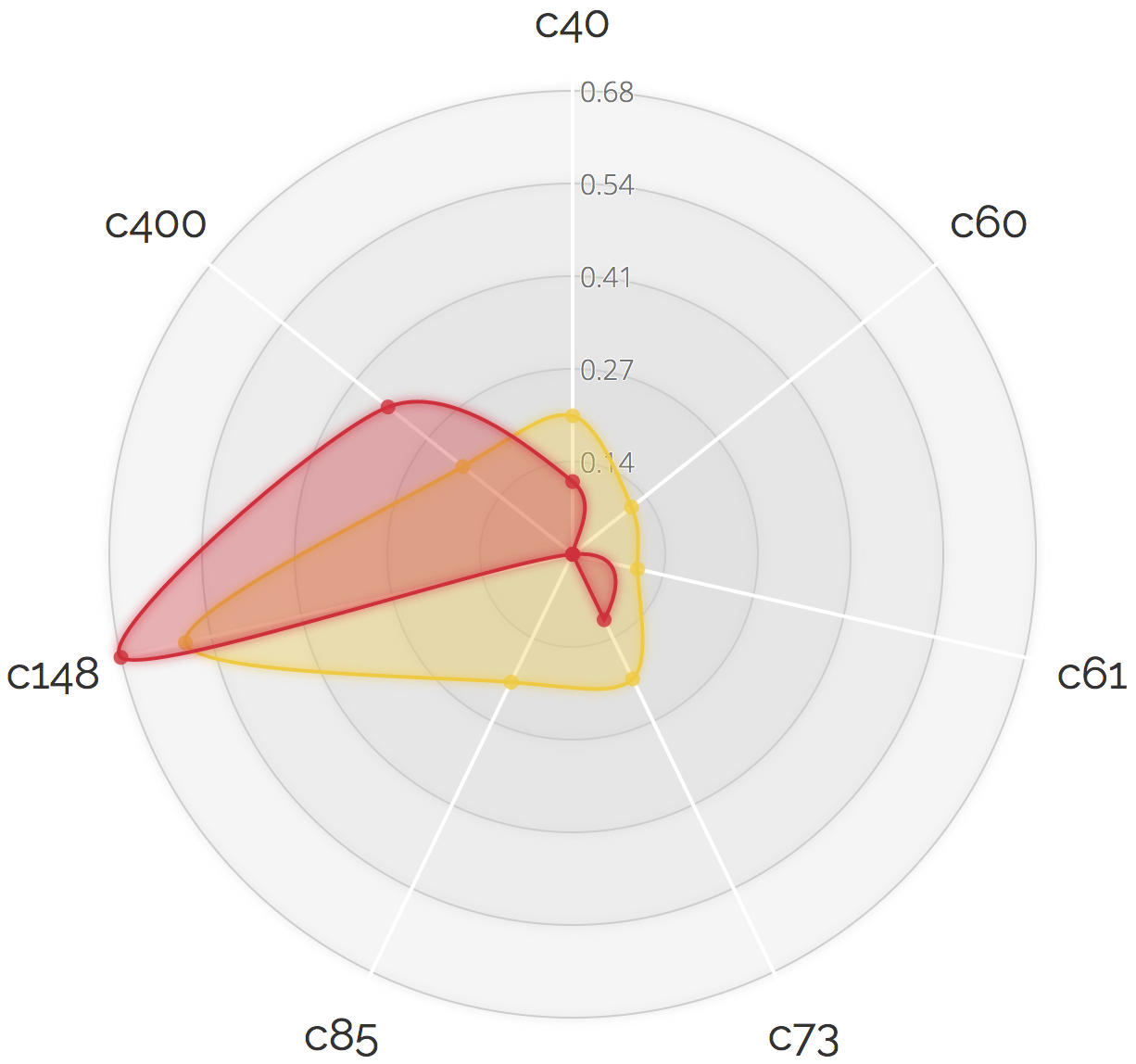
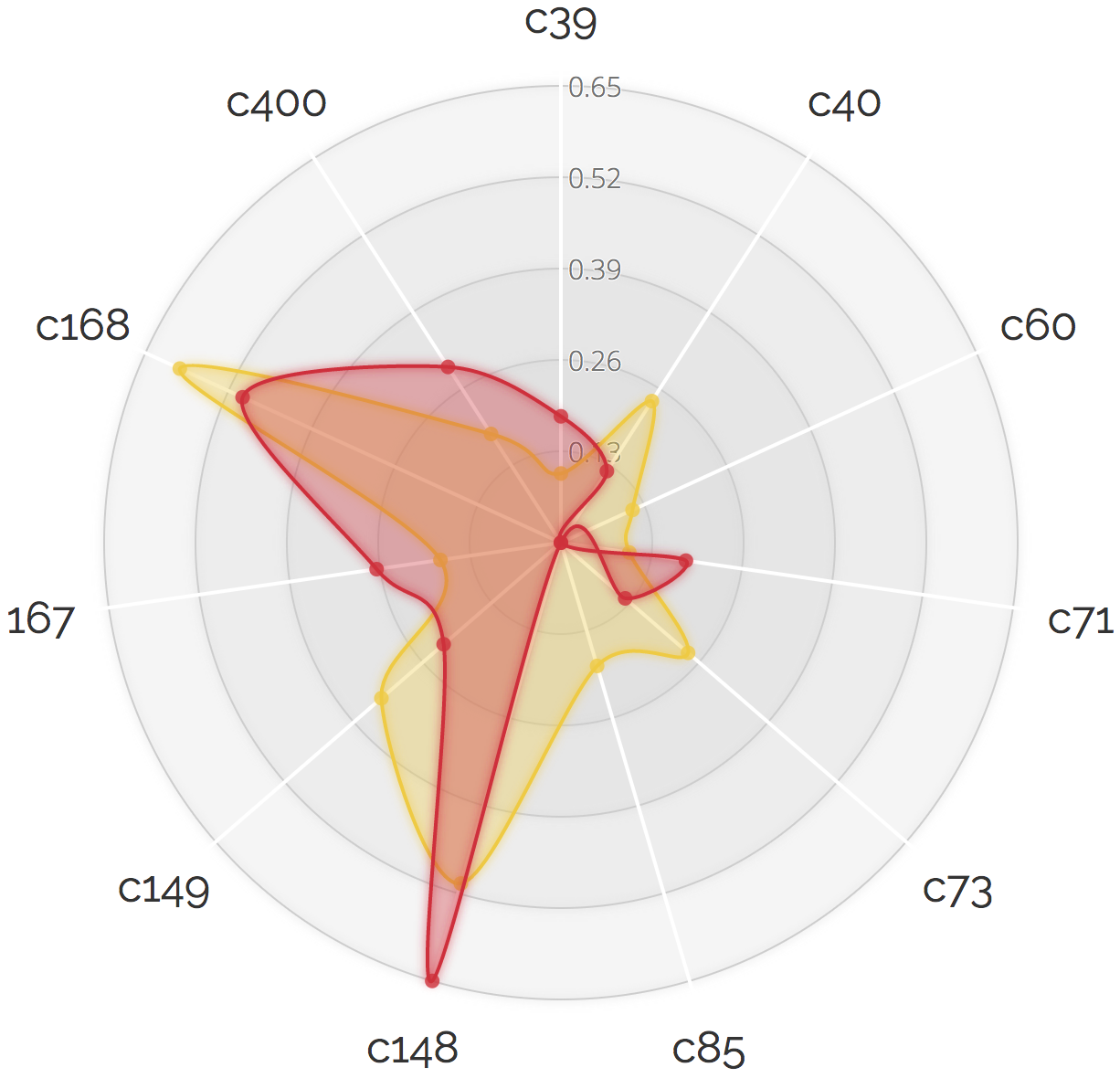
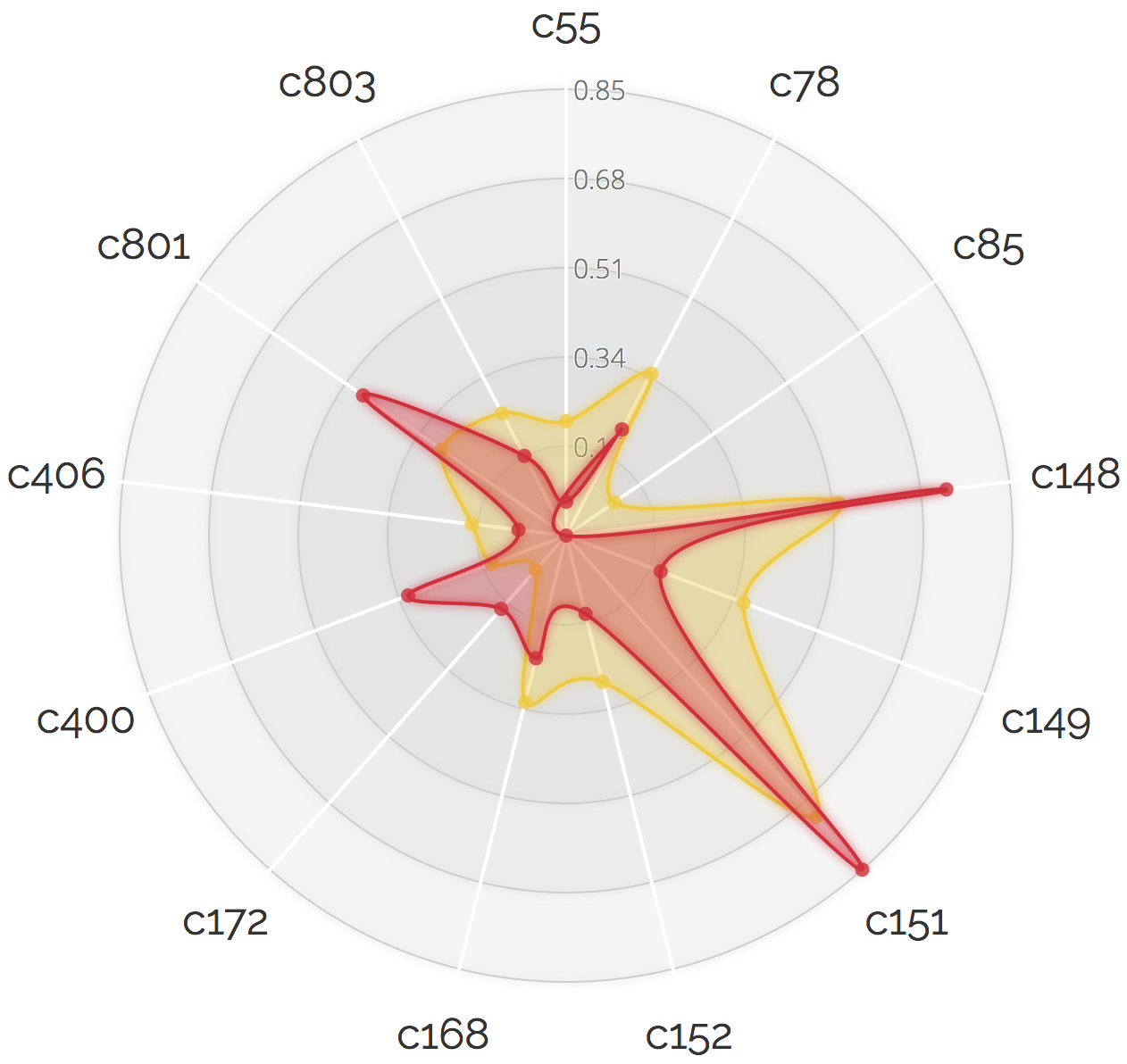
Several conclusions can be obtained after analyzing these two representations, almost at a glance. The clearest one (based on Fig. 6) is that the profiles of German and British repeaters are similar to each other them and very different from that of the Spanish. In the case of German and British repeaters, two groups of customers are important: clients without children (category c61 for Germany and category c73 for UK) and seniors with long stays (category c73 for both nationalities). In both cases, it is significant that they are single (category c149) and travel in pairs (category c150). With respect to the first-timers profile, and in the case of the British, the favorite destination is Mallorca (category c167). Several fields with unknown value are significant, such as civil status (category c148), customer segment (category c39), customer group (category c71), or habitual residence (category c400). Therefore, more information is required from this group of clients in order to be able to better establish their specific profile. However, Spanish repeaters belong to the young-client escape group (category c78), traveling without children (category c85), and with high frequency (category c55). It is also significant that they often are single (category c149) and male (category c1). Among Spanish, the first-timers profile is characterized by traveling as a family (category c151) and booking through an agency (category c801).







***Figure 6***. Chromosome proportions plots for Germany, UK, and Spain.

. 

*Figure 7*. Spiderweb representations for Germany (left), Spain (center) and UK (right). Repeaters and

first-timers profiles are represented by red and yellow areas, respectively. These areas are defined by the | Δ*pi* | for the most significant categories above threshold.

### 5. Discussion and Implications

The work in this paper allows us to scrutinize and present two different types of implications, namely, the methodological issues with respect to the application of the Big Data from CRM information, and those derived from the hotel chain study concerning the profile of the client. Both of them are summarized next after the discussion on the presented results.

5.1. Discussion on the Big-Case Study

In our study, the repeaters profile is more similar in German and British guests than in Spanish guests. In the case of foreign repeaters, they are characterized by being senior citizens traveling for long periods without children. The importance of this segment in the tourism market has already been pointed out, and its behavior has become of great interest due to its size and potential growth (Chen et al., 2013). Older consumers are presumed to have firmer brand loyalty than younger consumers (Henry, 2000). Also, German and British first-timers are significantly young clients on their holidays without children in our data, similar to the results by Gitelson & Crompton (1984), who noted that first-timers are more likely to be younger and single.

Spanish repeaters belong to the young-client escape group. These results differ for Spanish clients and match for German and British when compared with the results supported by Oppermann (1997), who demonstrated that first-timers tend to spend more money, but they stay for a shorter time than repeaters. Consistent findings were also observed by Wang (2004) and Lau & McKerchers (2004), however, contradictory findings regarding the length of stay were reported by Li et al. (2008), who concluded that first-timers are most likely to stay for longer periods, while repeaters are more likely to take weekend trips for visiting friends and relatives. The latter issue is easier for Spanish than for non-Spanish guests, since most of the chain establishments are located in Spain and this makes it more accessible to national customers to be able to move more frequently. This difference between the stay duration of Spanish and non-Spanish guests has been previously pointed out by Talón-Ballestero et al. (2016), who showed that the length of stay of foreign clients is longer in the Spanish hotel sector.

In addition, repeaters are single (specially for British and Spanish), in contrast with the ideas of Tasci (2016), who concluded that married clients were more loyal customers, and of Gitelson & Crompton (1984), who highlighted that the first-timers were mostly single. The fact of traveling without a family coincides in all three cases, which stands for a remarkable result, since according to Talón et al. (2016), up to 60% of the Spanish holiday market are families. Most of them are men and, in the case of the Spanish, they hired their last B2C visit. In making travel decisions, repeaters seem to rely more on their own experiences than on other information sources, hence, they spend much less time on planning (Li et al., 2008). On the other hand, Spanish first-timers have significantly purchased their last visit through an agency. According to Li et al. (2008), first-timers are more likely to rely on advice from travel professionals for making their travel decisions.

Women in the three nationalities are less often repeaters consistently in our data, which could be related to previous studies that have shown the different behavior in men and women in terms of their hotel preferences (Lutz & Ryan, 1993; McCleary et al. 1994; Sammons et al. 1999; Ariffin & Maghzi, 2012).

Interestingly, first-timers are less reported in the CRM systems, for instance in terms of their civil status, among others. This brings into scene the need for as complete as possible data recording in order to better identify customers and offering an adapted service, hence increasing their willingness to repeat (Padilla-Melendez & Garrido-Moreno, 2014).

5.2. Technical implications

Our proposal paves the way towards a systematic yet simple approach to exploit the data from CRM systems through Big Data in order to determine the client profiling. This analysis has also shown the remarkable effect of map-reduce techniques for yielding consistency to probability based statistical tests, thus making it evident that even simple statistical descriptions are more efficiently tackled than conventional and lower-scale approaches. To our understanding, the most relevant technical contribution of this work is to show that Big Data, with simple statistical concepts and structured data available in the organizations, yields a better knowledge of the client.

The data structure of CRMs is a strategic asset for companies and their information is confidential, hence, CRM access with research purposes can be possible through collaboration agreements between companies and universities with shared interest to address Big Data exploitation. This is a common situation in Data Science for companies, and in our work, it has allowed us to obtain useful information from the existing data structures in the organization. Nevertheless, the developed methodology can be completely replicated in different CRM databases by any practitioner or data scientist. In addition, our results have provided useful information to the managers of the hotel chain with a view to reconsider the data structure of the CRM and its possible modification, and hence to improve its exploitation.

In this work, the clustering developed by the chain has been used as a set of variables, but no additional client grouping has been generated by our study, because our objective was to show the possibilities of Big Data to exploit the information available in the CRM system by univariable study of categorical variables. There is no doubt that other advanced multivariable analytics can highlight interesting cross-relationships among the business-relevant variables identified here by simple proportion tests.

A detailed statistical study of convergence from the theoretical point of view has not been carried out. But still, in the analyzed data, the asymptotic behavior of the distributions has been empirically observed (see Figure 4) and this has been the expected one. A theoretical bootstrap convergence study with large data is a future and interesting work that exceeded our current objectives.

5.3. Hotel Chain Implications

Previous client classification studies based on statistical approaches have been used with varying success, but they will not always be capable of dealing with very large data sets. The use of Big Data techniques has allowed us to establish the repeaters vs first-timers profile in terms of a large number of features and categories. Our results show that nationality, gender, age, length of stay, family conditions or selling channel, are strongly relevant for this profiling in the studied hotel chain, and special attention should be paid to their correct recording in the CRM system.

The more information recorded in the CRM systems, the easier it will be to generate detailed profiles with simple Big Data techniques. Hence, hotel managers will improve their client knowledge for increasing satisfaction and loyalty, ensuring the repeating visit and the increased profit (Tseng & Wu, 2014). From a marketing viewpoint, these results can help to identify clusters of clients for guiding supply of the company products and services in terms of their needs.

Moreover, it would be desirable to scrutinize the reorientation towards the growing female segment (40,7% of the analyzed sample) and to establish different actions in terms of the selling channel (b2c for repeaters and agencies for first-timers). In addition, repeaters mostly travel without children, which can support the current chain reorientation towards non-family vacation segments and " adult only" products.

One of the limitations of our study comes from the consideration of a single hotel chain, and thus, of a specific CRM system. As pointed out, access to these data requires collaboration agreements with companies. Future research may use this methodology in different companies, because the method that can be easily applied to other CRM systems by Data Scientists and experts.

Not only the conceptual approach, but also the practical guidance for hospitality managers, can be usefully improved by our results. As it has been observed, the correct management through the Big Data techniques of the large volume of data generated by the clients in their stay, which are regularly collected in the CRM system, allows a greater and better knowledge of their characteristics, which will improve the customer satisfaction, carry out personalized marketing campaigns, as well as give offers to selected customers to book an adequate room at the selected rate.

Future research lines can generalize the current results, for instance, by comparing the differences between first-timers vs. repeaters for different types of Hotel Chains, different purposes of visit (business clients), and in different nationalities.

6. Final Conclusions

We concluded that the repeater profile in this chain corresponds to single, men, and traveling without children in the three scrutinized nationalities, however, there are differences among nationalities in terms of length of stay (larger in British and German than in Spanish) and age (senior in British and German vs. younger in Spanish). Moreover, due to the large number of tourists considered and the high volume of their handled information, the profile detected in this chain can be very useful not only to hotels, but also to tourist companies and to the tourist destinations themselves, in order to conveniently adapt their products and their marketing actions.

Overall, the great amount of available data from clients creates relevant opportunities for the hotel companies, which can turn into a strong competitive advantage. Further technical and more advanced tools will allow us to better exploit the best available information about the clients and their purchasing behavior. Our study has shown that even simple statistics as the proportion tests can be used for stating a solid large-scale information retrieval for client profiling, and it paves the way towards Big Data approaches yielding strong support for decision making of hotel managers.

**References**

Abu Kasim, N. A., & Minai, B. (2009). Linking CRM strategy, customer performance measures and performance in the hotel industry. *International Journal of Economics and Management*, *3*(2), 297-316.

Adomavicius, G., & Tuzhilin, A. (2001, 03). Using data mining methods to build customer profiles. *Computer,* *34*(3), 74-82. doi:10.1109/2.901170

Anwar, S. A., & Sohail, M. S. (2004, 04). Festival tourism in the United Arab Emirates: First-time versus repeat visitor perceptions. *Journal of Vacation Marketing,* *10*(2), 161-170. doi:10.1177/135676670401000206

Ariffin, A. A., & Maghzi, A. (2012, 03). A preliminary study on customer expectations of hotel hospitality: Influences of personal and hotel factors. *International Journal of Hospitality Management,* *31*(1), 191-198. doi:10.1016/j.ijhm.2011.04.012

Balomenou, N., & Garrod, B. (2014, 10). Using volunteer-employed photography to inform tourism planning decisions: A study of St David's Peninsula, Wales. Tourism Management, 44, 126-139. doi:10.1016/j.tourman.2014.02.015

Bayer, J. (2010, 09). Customer segmentation in the telecommunications industry. *Journal of Database Marketing & Customer Strategy Management,* *17*(3-4), 247-256. doi:10.1057/dbm.2010.21

Buttle, F. (2004). *Customer relationship management: Concepts and tools*. Amsterdam: Elsevier Butterworth-Heinemann.

Castellanos‐Verdugo, M., Oviedo‐García, M. D., Roldán, J. L., & Veerapermal, N. (2009, 04). The employee‐customer relationship quality. *International Journal of Contemporary Hospitality Management,* *21*(3), 251-274. doi:10.1108/09596110910948288

Chadha, A. (2015, 03). Case Study of Hotel Taj in the Context of CRM and Customer Retention. *KCAJBMR Kuwait Chapter of Arabian Journal of Business and Management Review,* *4*(7), 1-8. doi:10.12816/0018976

Chakravarti, D., & Day, G. S. (1991, 10). Market Driven Strategy: Processes for Creating Value. *Journal of Marketing,* *55*(4), 116. doi:10.2307/1251961

Chen, K., Liu, H., & Chang, F. (2013, 12). Essential customer service factors and the segmentation of older visitors within wellness tourism based on hot springs hotels. *International Journal of Hospitality Management,* *35*, 122-132. doi:10.1016/j.ijhm.2013.05.013

Dean, J., & Ghemawat, S. (2008, 01). MapReduce. *Communications of the ACM,* *51*(1), 107. doi:10.1145/1327452.1327492

Dev, C. S., & Olsen, M. D. (2000, 02). Marketing Challenges for the Next Decade. *Cornell Hotel and Restaurant Administration Quarterly,* *41*(1), 41-47. doi:10.1177/001088040004100122

Dimitriadis, S., & Stevens, E. (2008, 09). Integrated customer relationship management for service activities. *Managing Service Quality: An International Journal,* *18*(5), 496-511. doi:10.1108/09604520810898857

Dursun, A., & Caber, M. (2016, 04). Using data mining techniques for profiling profitable hotel customers: An application of RFM analysis. *Tourism Management Perspectives,* *18*, 153-160. doi:10.1016/j.tmp.2016.03.001

Efron, B., & Tibshirani, R. (1994). *An introduction to the bootstrap*. New York: Chapman & Hall.

Elfving, J., & Lemoine, K. (2012). Exploring the concept of Customer Relationship Management: emphasizing social. Phd Master thesis, Supervisor: Karin Brunsson. Department of Business Studies, Uppsala University, 2012-05-25

Fallon, P., & Schofield, P. (2003, 01). First-Timer Versus Repeat Visitor Satisfaction: The Case Of Orlando, Florida. *Tourism Analysis,* *8*(2), 205-210. doi:10.3727/108354203774076742

Fluker, M. R., & Turner, L. W. (2000, 05). Needs, Motivations, and Expectations of a Commercial Whitewater Rafting Experience. *Journal of Travel Research,* *38*(4), 380-389. doi:10.1177/004728750003800406

Formica, S., & Uysal, M. (1998, 04). Market Segmentation of an International Cultural-Historical Event in Italy. *Journal of Travel Research,* *36*(4), 16-24. doi:10.1177/004728759803600402

Fox, S., & Do, T. (2013, 09). Getting real about Big Data: Applying critical realism to analyse Big Data hype. *International Journal of Managing Projects in Business,* *6*(4), 739-760. doi:10.1108/ijmpb-08-2012-0049

George, G., Haas, M. R., & Pentland, A. (2014). Big Data and management. *Academy of Management Journal*, *57*(2), 321-326.

Gitelson, R. J., & Crompton, J. L. (1984, 01). Insights into the repeat vacation phenomenon. *Annals of Tourism Research,* *11*(2), 199-217. doi:10.1016/0160-7383(84)90070-7

Haynes, N. (2016). The evolution of competitor data collection in the hotel industry and its application to revenue management and pricing. *Journal of Revenue and Pricing Management*, *15*(3-4), 258-263.

Hendricks, K. B., Singhal, V. R., & Stratman, J. K. (2007). The impact of enterprise systems on corporate performance: A study of ERP, SCM, and CRM system implementations. *Journal* Henry, C. D. (2000, 07). Is customer loyalty a pernicious myth? *Business Horizons,* *43*(4), 13-16. doi:10.1016/s0007-6813(00)00066-5

Homburg, C., & Giering, A. (2000). Personal characteristics as moderators of the relationship between customer satisfaction and loyalty?an empirical analysis. *Psychology and Marketing,* *18*(1), 43-66. doi:10.1002/1520-6793(200101)18:13.0.co;2-i

Hjorth, L. (2012, 12). Relocating the mobile: A case study of locative media in Seoul, South Korea. Convergence: The International Journal of Research into New Media Technologies, 19(2), 237-249. doi:10.1177/1354856512462360

Katz, H. (2002). How to embrace CRM and make it succeed in an organization. *SYSPRO White Paper. SYSPRO, Costa Mesa, CA*.

Keramati, A., Mehrabi, H., & Mojir, N. (2010, 10). A process-oriented perspective on customer relationship management and organizational performance: An empirical investigation. *Industrial Marketing Management,* *39*(7), 1170-1185. doi:10.1016/j.indmarman.2010.02.001

Kim, B., & Choi, J. P. (2010, 06). Customer Information Sharing: Strategic Incentives and New Implications. *Journal of Economics & Management Strategy,* *19*(2), 403-433. doi:10.1111/j.1530-9134.2010.00256.x

Kim, C., & Lee, S. (2000, 07). Understanding the Cultural Differences in Tourist Motivation Between Anglo-American and Japanese Tourists. *Journal of Travel & Tourism Marketing,* *9*(1-2), 153-170. doi:10.1300/j073v09n01\_09

Kim, S. S., & Prideaux, B. (2005, 06). Marketing implications arising from a comparative study of international pleasure tourist motivations and other travel-related characteristics of visitors to Korea. *Tourism Management,* *26*(3), 347-357. doi:10.1016/j.tourman.2003.09.022

Kim, S., Jung, T., Suh, E., & Hwang, H. (2006, 07). Customer segmentation and strategy development based on customer lifetime value: A case study. *Expert Systems with Applications,* *31*(1), 101-107. doi:10.1016/j.eswa.2005.09.004

Kleiner, A., Talwalkar, A., Sarkar, P., & Jordan, M. I. (2014, 03). A scalable bootstrap for massive data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* *76*(4), 795-816. doi:10.1111/rssb.12050

Kotler, P. (2002). When to use CRM and When to forget it!. Paper presented at the Academy of Marketing Science, Sanibel Harbour Resort and Spa, 30 May.

Kozak, M. (2002, 06). Comparative analysis of tourist motivations by nationality and destinations. *Tourism Management,* *23*(3), 221-232. doi:10.1016/s0261-5177(01)00090-5

Lau, A. L., & Mckercher, B. (2004, 02). Exploration Versus Acquisition: A Comparison of First-Time and Repeat Visitors. *Journal of Travel Research,* *42*(3), 279-285. doi:10.1177/0047287503257502

Lee, S., Hwang, E., Jo, J. Y., & Kim, Y. (2016). Big Data Analysis with Hadoop on Personalized Incentive Model with Statistical Hotel Customer Data. *International Journal of Software Innovation*, *4*(3), 1-21.

Li, X., Cheng, C., Kim, H., & Petrick, J. F. (2008, 04). A systematic comparison of first-time and repeat visitors via a two-phase online survey. *Tourism Management,* *29*(2), 278-293. doi:10.1016/j.tourman.2007.03.010

Liang, F., Kim, J., & Song, Q. (2016, 07). A Bootstrap Metropolis–Hastings Algorithm for Bayesian Analysis of Big Data. *Technometrics,* *58*(3), 304-318. doi:10.1080/00401706.2016.1142905

Lin, Y., & Su, H. (2003, 08). Strategic analysis of customer relationship management—a field study on hotel enterprises. *Total Quality Management & Business Excellence,* *14*(6), 715-731. doi:10.1080/1478336032000053843

Liu, C. (2015). A Conceptual Framework of Analytical CRM in Big Data Age. *International Journal of Advanced Computer Science and Applications Ijacsa,* *6*(6). doi:10.14569/ijacsa.2015.060620

Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008, 06). Electronic word-of-mouth in hospitality and tourism management. Tourism Management, 29(3), 458-468. doi:10.1016/j.tourman.2007.05.011

Lutz, J., & Ryan, C. (1993, 10). Hotels and the businesswoman. *Tourism Management,* *14*(5), 349-356. doi:10.1016/0261-5177(93)90003-4

Mccleary, K. (1994, 04). Gender-based differences in business travelers' lodging preferences. *The Cornell Hotel and Restaurant Administration Quarterly,* *35*(2), 51-58. doi:10.1016/0010-8804(94)90019-1

Mckercher, B. (2004, 11). Understanding Tourism Behavior: Examining the Combined Effects of Prior Visitation History and Destination Status. *Journal of Travel Research,* *43*(2), 171-179. doi:10.1177/0047287504268246

Min, H., Min, H., & Emam, A. (2002, 11). A data mining approach to developing the profiles of hotel customers. *International Journal of Contemporary Hospitality Management,* *14*(6), 274-285. doi:10.1108/09596110210436814

Mok, C., & Armstrong, R. W. (1998, 10). Expectations for hotel service quality: Do they differ from culture to culture? *Journal of Vacation Marketing,* *4*(4), 381-391. doi:10.1177/135676679800400406

Morais, D. B., & Lin, C. (2010, 03). Why Do First-Time and Repeat Visitors Patronize a Destination? *Journal of Travel & Tourism Marketing,* *27*(2), 193-210. doi:10.1080/10548401003590443

Nasution, H. N., & Mavondo, F. T. (2008, 04). Organisational capabilities: Antecedents and implications for customer value. *European Journal of Marketing,* *42*(3/4), 477-501. doi:10.1108/03090560810853020

Nguyen, T. H., Sherif, J. S., & Newby, M. (2007, 05). Strategies for successful CRM implementation. *Information Management & Computer Security,* *15*(2), 102-115. doi:10.1108/09685220710748001

Noone, B. M. (2016). Pricing for hotel revenue management: Evolution in an era of price transparency. *Journal of Revenue and Pricing Management*, *15*(3-4), 264-269.

Oppermann, M. (1997, 05). First-time and repeat visitors to New Zealand. *Tourism Management,* *18*(3), 177-181. doi:10.1016/s0261-5177(96)00119-7

Özköse, H., Arı, E. S., & Gencer, C. (2015). Yesterday, today and tomorrow of Big Data. *Procedia-Social and Behavioral Sciences*, *195*, 1042-1050.

Padilla-Meléndez, A., & Garrido-Moreno, A. (2013, 06). Customer relationship management in hotels: Examining critical success factors. *Current Issues in Tourism,* *17*(5), 387-396. doi:10.1080/13683500.2013.805734

Pan, B., & Li, X. R. (2011). The long tail of destination image and online marketing. Annals of Tourism Research, 38(1), 132-152.

Pan, B. I. N. G., & Yang, Y. A. N. G. (2016). Monitoring and forecasting tourist activities with Big Data. *Management Science in Hospitality and Tourism: Theory, Practice, and Applications*, *43*.

Papastathopoulou, P., Avlonitis, G. J., & Panagopoulos, N. G. (2007, 04). Intraorganizational information and communication technology diffusion: Implications for industrial sellers and buyers. *Industrial Marketing Management,* *36*(3), 322-336. doi:10.1016/j.indmarman.2005.10.002

Payne, A., & Frow, P. (2005, 10). A Strategic Framework for Customer Relationship Management. *Journal of Marketing,* *69*(4), 167-176. doi:10.1509/jmkg.2005.69.4.167

Peter, T. Use hotel data to drive growth. http://www.hotelnewsnow.com/Article/14553/ Use-hotel-data-to-drive-growth, accessed 26/01/2016.

Petrick, J. F. (2004, 08). Are loyal visitors desired visitors? *Tourism Management,* *25*(4), 463-470. doi:10.1016/s0261-5177(03)00116-x

Petrick, J. F. (2005, 10). Segmenting cruise passengers with price sensitivity. *Tourism Management,* *26*(5), 753-762. doi:10.1016/j.tourman.2004.03.015

Ramos, C. M., Martins, D. J., Serra, F., Lam, R., Cardoso, P. J., Correia, M. B., & Rodrigues, J. M. (2017). Framework for a Hospitality Big Data Warehouse: The Implementation of an Efficient Hospitality Business Intelligence System. *International Journal of Information Systems in the Service Sector*, *9*(2), 27-45.

Rong, J., Vu, H. Q., Law, R., & Li, G. (2012, 08). A behavioral analysis of web sharers and browsers in Hong Kong using targeted association rule mining. Tourism Management, 33(4), 731-740. doi:10.1016/j.tourman.2011.08.006

Rosenberger, M. (2015). Social Customer Relationship Management: An Architectural Exploration of the Components. *Open and Big Data Management and Innovation Lecture Notes in Computer Science,* 372-385. doi:10.1007/978-3-319-25013-7\_30

Ryu, K., & Han, H. (2011, 09). New or repeat customers: How does physical environment influence their restaurant experience? *International Journal of Hospitality Management,* *30*(3), 599-611. doi:10.1016/j.ijhm.2010.11.004

Sá, J. P. Marques De (2003). *Applied Statistics: Using SPSS, STATISTICA, and MATLAB*. Berlin: Springer.

Sammons, G., Moreo, P., Benson, L. F., & Demicco, F. (1999, 05). Analysis of Female Business Travelers' Selection of Lodging Accommodations. *Journal of Travel & Tourism Marketing,* *8*(1), 65-83. doi:10.1300/j073v08n01\_04

Santana-Cerdeña, L. Ramos, S., Bobur S. (2014) “Potencial y Retos del Big Data en Turismo” in *X Congreso de Turismo y Tecnologías de la Información y las Comunicaciones, Universidad de Málaga, Facultad de Turismo, pp.21-35*

Sarmaniotis, C., Assimakopoulos, C., & Papaioannou, E. (2013, 07). Successful implementation of CRM in luxury hotels: Determinants and measurements. *EuroMed Journal of Business,* *8*(2), 134-153. doi:10.1108/emjb-06-2013-0031

Seddighi, H., Nuttall, M., & Theocharous, A. (2001, 04). Does cultural background of tourists influence the destination choice? an empirical study with special reference to political instability. *Tourism Management,* *22*(2), 181-191. doi:10.1016/s0261-5177(00)00046-7

Shaw, M. J., Subramaniam, C., Tan, G. W., & Welge, M. E. (2001, 05). Knowledge management and data mining for marketing. *Decision Support Systems,* *31*(1), 127-137. doi:10.1016/s0167-9236(00)00123-8

Sigala, M. (2005, 09). Integrating customer relationship management in hotel operations: Managerial and operational implications. *International Journal of Hospitality Management,* *24*(3), 391-413. doi:10.1016/j.ijhm.2004.08.008

Silva, A. P., & Mateus, G. R. (2003, 05). A Location-Based Service Application for a Mobile Computing Environment. Simulation, 79(5-6), 343-360. doi:10.1177/0037549703037151

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017, 01). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research,* *70*, 263-286. doi:10.1016/j.jbusres.2016.08.001

Skogland, I., & Siguaw, J. A. (2004, 08). Are Your Satisfied Customers Loyal? *Cornell Hotel and Restaurant Administration Quarterly,* *45*(3), 221-234. doi:10.1177/0010880404265231

Soguero-Ruiz, C., & Jenssen, R. (2015, 09). Kernel covariance series smoothing. *IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)*. doi:10.1109/mlsp.2015.7324380

Soltani, Z., & Navimipour, N. J. (2016, 08). Customer relationship management mechanisms: A systematic review of the state of the art literature and recommendations for future research. *Computers in Human Behavior,* *61*, 667-688. doi:10.1016/j.chb.2016.03.008

Song, H., & Liu, H. (2017). Predicting Tourist Demand Using Big Data. In *Analytics in Smart Tourism Design* (pp. 13-29). Springer International Publishing.

Talón-Ballestero, P., González-Serrano, L. & Rodríguez-Antón, J.M. (2016). *Fundamentos de Dirección Hotelera.* Volumen I. Madrid: Síntesis. ISBN: 978-84-9077-392-5

Tasci, A. D. (2016). A quest for destination loyalty by profiling loyal travelers. *Journal of Destination Marketing & Management*. http://dx.doi.org/10.1016/j.jdmm.2016.04.001

Tohidi, H., & Jabbari, M. M. (2012). CRM in Organizational Structure Design. *Procedia Technology,* *1*, 579-582. doi:10.1016/j.protcy.2012.02.126

Tseng, S., & Wu, P. (2014, 03). The impact of customer knowledge and customer relationship management on service quality. *International Journal of Quality and Service Sciences,* *6*(1), 77-96. doi:10.1108/ijqss-08-2012-0014

Tseng, C., Wu, B., Morrison, A. M., Zhang, J., & Chen, Y. (2015, 02). Travel blogs on China as a destination image formation agent: A qualitative analysis using Leximancer. Tourism Management, 46, 347-358. doi:10.1016/j.tourman.2014.07.012

Verdugo, C.M., Oviedo-Garcia, A.M., Roldan, L.J. (2009), The employee-customer relationship quality: Antecedents and consequences in the hotel industry. International Journal of Contemporary Hospitality Management, 21(3), 251-274.

Vu, H. Q., Li, G., Law, R., & Ye, B. H. (2015, 02). Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos. *Tourism Management,* *46*, 222-232. doi:10.1016/j.tourman.2014.07.003

Wang, D. (2004, 02). Tourist Behaviour and Repeat  Visitation to Hong Kong. *Tourism Geographies,* *6*(1), 99-118. doi:10.1080/14616680320001722355

Wang, W. (2013, 12). A note on bootstrap confidence intervals for proportions. *Statistics & Probability Letters,* *83*(12), 2699-2702. doi:10.1016/j.spl.2013.09.005

Wind, Y. (1978, 08). Issues and Advances in Segmentation Research. *Journal of Marketing Research,* *15*(3), 317. doi:10.2307/3150580

Wind, Y., & Lerner, D. (1979, 02). On the Measurement of Purchase Data: Surveys versus Purchase Diaries. *Journal of Marketing Research,* *16*(1), 39. doi:10.2307/3150872

Woodside, A. G. (2000). *Consumer pshychology of tourism, hospitality, and leisure*. Wallingford, Oxon, UK: CABI Pub.

Wu, S., & Li, P. (2011, 06). The relationships between CRM, RQ, and CLV based on different hotel preferences. *International Journal of Hospitality Management,* *30*(2), 262-271. doi:10.1016/j.ijhm.2010.09.011

Wu, S., & Lu, C. (2012, 03). The relationship between CRM, RM, and business performance: A study of the hotel industry in Taiwan. *International Journal of Hospitality Management,* *31*(1), 276-285. doi:10.1016/j.ijhm.2011.06.012

Xiang, Z., Schwartz, Z., Gerdes, J. H., & Uysal, M. (2015). What can Big Data and text analytics tell us about hotel guest experience and satisfaction?. *International Journal of Hospitality Management*, *44*, 120-130.

Yang, Y., Pan, B., & Song, H. (2014). Predicting hotel demand using destination marketing organization’s web traffic data. *Journal of Travel Research*, *53*(4), 433-447.

Ye, Q., Law, R., & Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management*, *28*(1), 180-182.

Yoo, M., & Bai, B. (2013, 06). Customer loyalty marketing research: A comparative approach between hospitality and business journals. *International Journal of Hospitality Management,* *33*, 166-177. doi:10.1016/j.ijhm.2012.07.009

Zhang, Y., Shu, S., Ji, Z., & Wang, Y. (2015, March). A Study of the Commercial Application of Big Data of the International Hotel Group in China: Based on the Case Study of Marriott International. In *Big Data Computing Service and Applications* (pp. 412-417).