

Segmenting visitors of cultural events: the case of Christmas Market

Juan Gabriel Brida; Marta Disegna*; Raffaele Scuderi

School of Economics and Management, Free University of Bozen-Bolzano Universitätsplatz 1 - piazza
Università, 1 39100 Bozen-Bolzano – Italy

Abstract

Market segmentation in tourism makes use of sets of powerful analytical tools for the sake of planning and managing demand-oriented policies. This paper contributes to this strand of literature by segmenting tourists visiting a cultural event. We utilize the Bagged Clustering method, a combination of traditional partitioning and hierarchical techniques, which is proven to be more effective. An ad-hoc survey was conducted in 2011 among the Italian visitors of the Christmas Market in Merano, Northern Italy. A total of 802 questionnaires were collected. In discussing the results, marketing and managerial implications are stressed for both policymakers and local organizers.

Keywords: Bagged Clustering, Logit model, cultural event, Christmas Market, market segmentation

* Corresponding author. Address: School of Economics and Management, Free University of Bolzano, Piazzetta dell'Università, 39031 Brunico (BZ), Italy. Tel.: +39 0471 013293; fax: +39 0471 013009.

E-mail addresses: JuanGabriel.Brida@unibz.it (J.G. Brida), marta.disegna@unibz.it (M. Disegna), Raffaele.Scuderi@unibz.it (R. Scuderi).

1. Introduction

Christmas markets (CMs) have become important yearly events for many cities in Italy, in particular of the South Tyrol region. The perception of the Advent atmosphere, as well as of markets as “authentic” events of the region (Brida et al., 2013a, 2012a), attract significant flows of visitors. Nowadays Christmas fairs are held in five cities of South Tyrol during the four weeks of Advent, from the end of November till Christmas Eve. They attract a large number of tourists coming mainly from Italy and often visiting the “circuit” of South Tyrol markets of Bolzano, Merano, Bressanone, Brunico, and Vipiteno. The extension of the opening season for the majority of hotels testifies to the importance of this event for the local economy, confirming that “events may be of especial interest in terms of reducing seasonality” (Nicholson & Pearce, 2000).

Understanding desires, perceptions, characteristics, and needs of visitors in a place where CM has such a deep impact can be crucial for local policymakers in order to address their actions and shape their policies according to the characteristics of the demand. For a long time visitors to cultural attractions and events have been treated as homogeneous mass of people. Instead, recent tourist literature tends to consider them as heterogeneous market with different characteristics, perceptions, and needs (Hughes, 2002). Consistently with this view, the goal of this study is to perform a segmentation analysis of CM visitors. The Bagged Clustering (BC) technique (Leisch, 1999), a segmentation method that is not common in the marketing and tourism field, was adopted in this study. In particular, two objectives are pursued. First, the paper aims at finding homogeneous groups of visitors according to their reasons for visiting the CM. Then clusters of tourists with similar levels of travel expenditure are analysed. The BC algorithm combines sequentially partitioning and hierarchical clustering methods presenting several advantages compared to classic unsupervised techniques. In particular, this method allows obtaining more robust and stable segment solution than other classic methods thanks to the adoption of the bagging (“bootstrap aggregating”) procedure (Breiman, 1996). The adopted methodology presents several advantages with respect to classical techniques, such as the improved stability of results, a reduced dependency on the starting solution, and the possibility of using large datasets. This implies that managerial implications can be drawn based on more robust empirical evidence. BC has been successfully applied by its author or his research team, for the sake of tourism market segmentation (Dolnicar et al., 2008; Dolnicar & Leisch, 2003, 2000), but it has been applied infrequently by other scholars in

tourism. In particular, to the best knowledge of the authors this work is the first attempt to use BC in a cultural event analysis. Data were collected from an ad-hoc survey conducted in 2011 during the four weeks of advent (from 30 November to 24 December) among the Italian visitors of the CM of Merano, Northern Italy.

The paper is organized as follows. In Section 2, a literature review on the applied segmentation studies is provided. In particular, some key studies of segmentation in tourism and cultural events are reviewed. In Section 3 an insight on the methodology is given, focusing on both the questionnaire design and the BC method employed to run the empirical investigation. In Section 4, results emerging from the cluster analysis are reported. Conclusions, limitations of the study, and practical implications are summarized in the last section.

2. Literature Review and Background

2.1 Market segmentation in tourism

Market segmentation is a simple concept that has been considered widely as a key instrument in the field of strategic marketing (Kruger et al., 2011; Najmi et al., 2010). It consists of subdividing a market into smaller and homogeneous groups. It assumes that markets and the involved individuals are not homogeneous and therefore, no single supply will satisfy everyone (see among others Kruger et al., 2011; Najmi et al., 2010; Tkaczynski & Rudle-Thiele, 2010; Kotler & Armstrong, 1999). In general, this allows the marketers to do direct efforts towards the groups of customers that are resulting more economically significant (Thompson & Schofield, 2009). Tourism is a natural extension of the market segmentation analysis. According to Lee and Beeler (2009), and Koc and Altinay (2007), developing and sustaining competitive advantage in competitive tourism markets largely depends upon on how well visitors are known. Kau and Lim (2005) stress that market segmentation allow to destination planners “to allocate resources more effectively in attracting distinct and unique groups of travellers”.

Since the introduction of market segmentation in the late 1950s, the number and type of approaches has grown enormously (Dolnicar & Leisch, 2004; Liao et al., 2012). This has happened also in the tourism literature. For an extensive review of recent academic literature about market segmentation in tourism see Cohen et al. (2014) and Pesonen (2013). The two

major approaches for segmenting individuals are *a priori* and *a posteriori*. The first approach consists in identifying the groups using a predefined criterion that is expected to cause heterogeneity among the customers. In the second approach, groups are recognized by applying quantitative method of data analysis. Among *a posteriori* segmentation approaches, cluster analysis remains the most popular method and the most frequently used in the literature (Jain, 2010; Dolnicar, 2002; Wedel & Kamakura, 2000). Clustering methods are generally divided into three categories: non-overlapping algorithms (each object is part of a single segment – Tuma et al., 2011); overlapping algorithms (an object may belong to more than one cluster – Wedel & Kamakura, 2000); fuzzy algorithms (each object is assigned by a degree of membership to a segment – Franke et al., 2009; Tuma et al., 2011). Hierarchical (agglomerative) and non-hierarchical (partitioning) methods are two common approaches that can be classified within non-overlapping algorithms. They have been used in the marketing and tourism literature very frequently (Tuma et al., 2011; Dolnicar & Leisch, 2004; Dolnicar, 2002, 2003).

Hierarchical methods aim to find clusters by iteratively joining the “closest” clusters composed of one or more observations (agglomerative clustering), or splitting the “furthest” clusters (divisive clustering). Standard partitioning methods group the observations around a centre in order to find a segmentation of the set of units in a fixed number of clusters, decided *a priori*. In marketing and tourism studies, the most commonly used algorithm in this category is the *k*-means (Dolnicar, 2002; Tuma et al., 2011). In general, partitioning methods are more flexible and perform better with large dataset than hierarchical methods (Dolnicar, 2002; Dolnicar & Leisch, 2004; Everitt et al., 2011; Brida et al., 2013b). The latter have the disadvantage that once observations are merged with others in a group, they cannot be removed from that cluster. Therefore, application of hierarchical methods is not always justified in market segmentation given that it presupposes an underlying hierarchy among the objects or respondents to be clustered (Tuma et al., 2011).

K-means instead strongly depends on the starting selected centres because it is based on iterative stochastic procedures. Running *k*-means algorithm twice on the same dataset with different starting centres may result in two different solutions: the less clear the hidden data structure, the higher the difference between two solutions. This causes *k*-means to be an unstable algorithm though widely used. The reason is related to the absence of a global solution and to the fact that at each iteration one may only find a local one. Another important disadvantage that occur when using *k*-means is that the number of clusters has to be selected in advance. In tourism studies using non-hierarchical algorithms, it is a common practice to

decide the number of groups on the basis of practical and subjective preference (Choi, 2011; Konu et al., 2011; Albalade & Bel, 2010; Pérez & Nadal, 2005) or derive this information from applying a hierarchical cluster method (Claver-Cortés et al., 2007; Bigné & Andreu, 2004; Chen & Hsu, 1999; Punj & Steward, 1983). Although many internal validity indices were developed in order to drive the researcher to select this number properly (see for example Handl et al., 2005), none has yet been globally accepted and in the tourism field they have not been widely applied (see Brida et al., 2012b for an example of their application).

Operationally, researchers can choose among a great number of clustering methods and each of them may conduct to a peculiar description of the data. This implies that “different methods present different views of data” (Leisch, 2006). Unfortunately, as emphasized by many researchers, no absolutely “correct” or commonly shared way to segment a market exists in the literature (Brida et al., 2012b; Tkaczynski & Rundle-Thiele, 2011; Kotler et al., 2010; Dolnicar et al., 2008; Beane & Ennis, 1987).

Recently, Leisch (1999) proposed the use of the BC, a method that combines sequentially partitioning and hierarchical clustering methods. It can be seen as both a complexity-reducing pre-processing stage for the hierarchical methods, and a procedure combining several results from partitioning (Kang et al., 2008; Leisch, 1999). It performs better in comparison to other standard clustering methods for both continuous and binary data sets, and it overcomes many limitations of either partitioning and hierarchical algorithms (Leisch, 1999). In particular, the main advantage of this algorithm, with respect to a partitioning one, is that it is not necessary to impose the number of clusters *a priori*. In addition, Dolnicar and Leisch (2004) showed that BC could provide more stable results than other classic methods. In fact, the bagging (“bootstrap aggregating”) procedure (Breiman, 1996) adopted in the BC algorithm is a resampling method that aims to improve the accuracy of results from unstable procedures.

2.2 Segmentation of visitors to cultural events

In this study we focus on the segmentation analysis of events, which has been the object of many studies. The comprehensive review of Tkaczynski and Rundle-Thiele (2011) reports that many different kinds of techniques have been applied. Papers making use of cluster analysis have been in a limited number, but the type of events under investigation is wide. These works have in common the goal of investigating the extent to which each segment is attracted by the cultural event or local tourist attractions, or rather specific motivations or socioeconomic characteristics are proper of it. For instance, different levels of cultural

orientation (i.e., low, medium, high) are reported by Lee and Lee (2001), who classify visitors according to their motivations to attend a cultural festival. Prentice and Andersen (2003) find seven clusters of visitors to Edinburgh festival according to consumption styles: serious consumers of international culture, British drama-going socializers, Scots performing arts attenders, Scottish experience tourists, gallery-goers, incidental festival-goers, accidental festival-goers. Lee et al. (2004) report four groups of visitors at the Kyongju World Culture Expo, South Korea, characterized by seeking culture and family, multi-purpose, escape, event. Chang (2006) stresses that the motivations of visitors to an aboriginal cultural festival in Taiwan can be grouped into equilibrium recovery, participation and learning, novelty-seeking, socialization, cultural exploration. De Guzman et al. (2006) investigate motivational factors at the WOW Philippines tourist festival, and find four groups of, respectively, “binding, bonding, blazing and bracing” tourists. Li et al. (2009) explore motivations of a rural festival attendees, and label their five clusters as family travellers, event enthusiasts, loyal festival goers, escapers, social gathering lovers. Oom do Valle et al. (2010) study the attendees’ profiles of a set of different types of events held in Faro, Portugal and detect three segments of tourists with different levels of satisfaction. The only paper investigating CMs is by Brida et al. (2013a), who survey visitors in three Italian cities and find three segments they name business people, Christmas fans, general tourists.

3. Methodology

3.1 Data and structure of the questionnaire

The research is based on a survey conducted in 2011 during the four weeks of Advent (November 30 – December 24). The convenience sampling method (Cochran, 1977) was adopted assuming 98% level confidence, $\pm 4\%$ margin of error, and 0.5 proportion, for a conservative estimate of the sample size, with a minimum sample size of 848. An overall number of 802 Italian visitors to the Merano CM were successfully interviewed, both tourists and same day-visitors, whereas local residents were excluded. This sample size corresponds to the assumption that the level of confidence is equal to 97%, remaining the other parameters unchanged.

The majority (62%) were interviewed at the end or during their visit. Interviewees came mainly from neighbouring regions of the Centre and East territories of Northern Italy: 25.1% from Lombardia, 19.7% from Veneto, 14.8% from Emilia-Romagna, and 9.73% from Trentino-Alto Adige. Interviews were held in the most visited parts of the CM, mainly (86%)

during week-end days (from Friday to Sunday). Most of the interviews (52%) were conducted in the afternoon or evening and all of them were collected during period of good weather. In order to encourage cooperative behaviour, respondents were informed that the research had exclusively scientific aims, and that impartiality in the data analysis was guaranteed. Furthermore a pilot survey was carried out to test the questionnaire before conducting the full survey, in order to avoid biases related to its structure and wording.

The questionnaires were anonymous and self-administered in three languages (Italian, German and English) and a research team member was present in order to solve the questions or doubts that emerged among interviewers. Interviewers selected only one person per household or travel party that passed through a previously selected spot. This sampling method was chosen because no sufficient information on the characteristics of visitors of the CM existed in order to apply a probabilistic design. Table 1 resumes the questionnaire sections schematically.

Table 1. Structure of the questionnaire.

Sections	Description	Categories of variables
I	Christmas Market information	Repeat visiting; number of Christmas Markets visited in the last 5 years; travel party characteristics; factors that stimulated the visit [*] ; shopping expenditure at the Christmas Market; authenticity perception [*] .
II	Trip information	Motives of the trip [*] ; number of nights; travel expenditure (accommodation, food and beverage, shopping) per person per day; distance covered to reach Merano.
III	Interviewees' profile	Socio-demographic and economic characteristics of interviewees and their families.

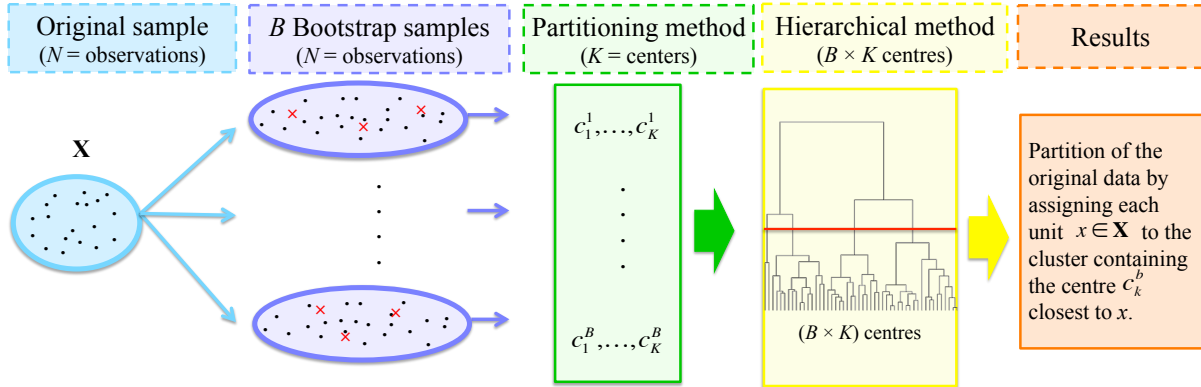
Notes: ^{*} A 5-points Likert scale was used.

3.2 Bagged clustering

Figure 1 schematically shows the steps of the BC method as proposed by Leisch (1999). Let us assume that \mathbf{X} is the initial dataset of size N , on which B bootstrap samples are drawn with replacement. A partitioning method, called “base” method, is chosen by the researcher (e.g. k -means) and applied to each bootstrapped sample. This creates $(B \times K)$ centres c_k^b , where K is the selected number of centres in a classic k -means algorithm in order to find partitions of units $k = 1, \dots, K$; and $b = 1, \dots, B$ is the bootstrap replication. The $(B \times K)$ centers are combined in a new dataset $\mathbf{C}_{B \times K}$ on which a hierarchical clustering method is run. The result is represented through a dendrogram, whose investigation allows obtaining the

“best” centers partition. Finally, each statistical unit is assigned to its closest centre, and consequently to a cluster. This way, also the “best” partition of the original units is obtained.

Figure 1. Bagged Clustering algorithm.



BC is more stable of k -means algorithm as it has a less strong dependence on the starting solution. Running the base method (i.e. the partitioning method chosen) on B bootstrapped sample is in fact equivalent to running B times the base method, and then obtaining the final solution that properly summarizes all these results. The BC overcomes the issue of selecting the number of groups. Although an initial choice of K is required, it does not affect the final results. In fact, the “final” number of clusters is obtained *a posteriori* as a result of the hierarchical algorithm. Still compared to k -means, which tends to identify often unrealistic equally-sized clusters, BC can reveal also unequally sized groupings and it is particularly suitable for detecting niche segments (Dolnicar & Leisch, 2004).

In this study, the BC algorithm was applied to two separate sets of segmentation variables. At first tourists were grouped according to their motivations by using variables related to push factors for visiting the CM. The objective was to find “motivations clusters”, that is homogeneous groups of people from the point of view of their motivations in visiting the CM. This clustering analysis could be of particular importance in order to distinguish attitudes that would not necessarily be related to the search for the Christmas atmosphere, which is commonly believed to be the main motive of the visit to a CM. The second criterion to choose the segmentation variables was based on the behaviour of visitors with respect to individual travel spending (“expenditures clusters”). This second analysis aims to find groups of visitors with similar levels of expenditure in different travel items, such as the purchase of products at CM and in other shops in the city, accommodation, food and beverage, transportation. This

analysis can indicate the presence of groups of tourists whose common attitudes may result in a similar level of expenditure, and thus in a similar economic impact on the territory.

The distance between segmentation variables was computed by using the Manhattan distance regarding the motivations, whereas in expenditure levels it was measured via the Euclidean distance. These parameters were chosen because they provided for the best performances in previous studies which used simulated artificial data sets with similar characteristics to the one used in this paper (Dolnicar & Leisch, 2004, 2000). The base method for clustering was the *k*-means, with $K=20$ centres and 10,000 iterations. The technique was applied to $B=50$ bootstrapped samples, which overall created $B \times K=1000$ centres which were then hierarchically clustered using Ward's agglomerative linkage method. In order to find the "best" partition we calculated the aggregation distances between two consecutive partitions (the higher the distance the higher the difference among the clusters), and three internal validity measures commonly used in the literature: Dunn Index (Dunn, 1974), Davies-Bouldin Index (Davies & Bouldin, 1979), and Silhouette index (Rousseeuw, 1987). It is important to underline that in practice the value of these internal validity measures must be interpreted as a guideline rather than absolute criterion (Vesanto & Alhoniemi, 2000). Finally, in the profiling stage of the two segmentation analysis, selected clusters are examined more in depth via regressions (in particular, binary and multinomial logistic regressions were adopted). The goal is to provide policy indications on ways to improve the visibility and impact of the event on specific groups.

4. Results

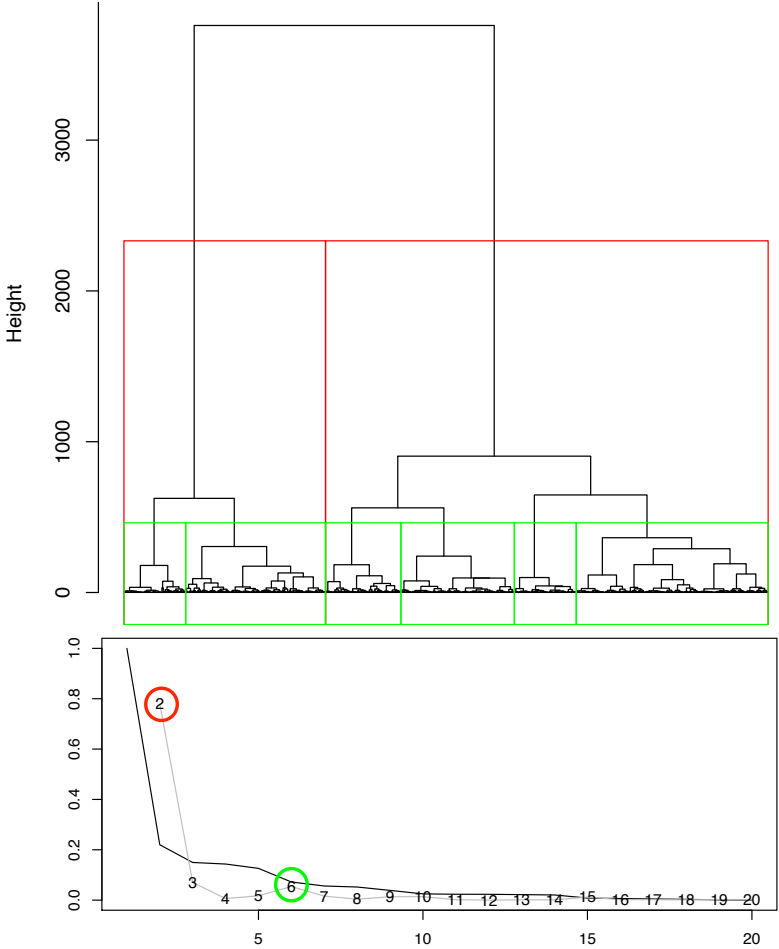
Results of the BC method for the "motivations clusters" analysis are reported by Figures 2 and 3, whereas Figures 4 and 5 display those of the "expenditures clusters" analysis.

The top parts of the Figures 2 and 4 report dendrograms that derived from the hierarchical clustering method applied to the centres. The plots under the dendrograms show aggregation distances of each of them, where the black line reports standardized absolute heights and the grey one stands for first differences. "Sudden bends" of the black line and/or local peaks of the grey one drive the selection of the number of clusters. The peak in the grey line (Figure 2) suggests that for the "motivations clusters" the correct number of groups is two, followed by the six-clusters solution. Results of the "expenditures clusters" (Figure 4) suggest instead that the two and six-clusters solutions are almost equally good solutions. Table 2 presents the results of the three selected internal validity measures calculated for all cluster partitions from

2 to 10. For “motivations clusters” the two clusters partition is selected as the best option for all the indexes used. Furthermore, the Davies-Bouldin index presents a local minimum also for the six-clusters solution, confirming the result obtained from the check of the aggregate distances. In contrast, the three internal validity measures do not identify the “best” partition for the “expenditures clusters”: Silhouette index increases as the number of clusters; Dunn index find a maximum for the nine-clusters solution (but the differences in value with the 6, 7, and 8 clusters solutions are very small); Davies-Bouldin index decreases as the number of clusters.

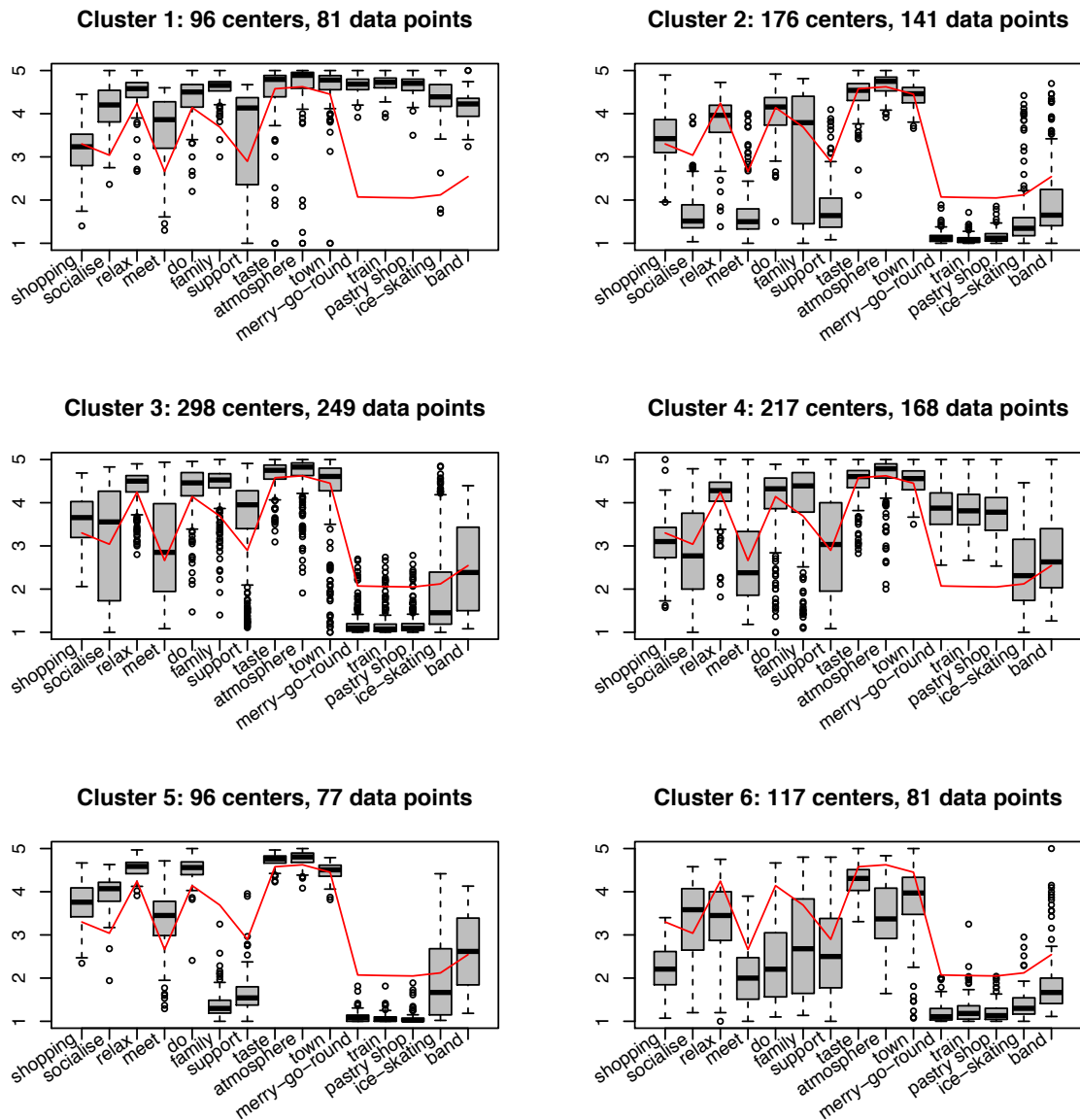
Either for the “motivations clusters” and the “expenditures clusters”, the two-clusters solution divides the sample into two large groups that give little information regarding visitors’ behaviour. For this reason, in the following section, the six clusters solution (the second best option in both analysis) is discussed for both motivations and expenditures.

Figure 2. Motivations clusters: Bagged Clustering dendrogram together with the plot regarding the relative height of aggregation (black line) and the first differences (grey line).



However, it is easy to obtain information regarding the two-clusters solution from the six-clusters one. Concerning “motivations clusters”, cluster 1 (two-clusters) is the sum of cluster 1 and 4 (six-clusters), whereas cluster 2 (two-clusters) is the sum of the remainder (six-clusters). For “expenditures clusters”, cluster 1 (two-clusters) is the sum of cluster 1 and 2 (six-clusters) while cluster 2 (two-clusters) is the sum of the remaining clusters (six-clusters).

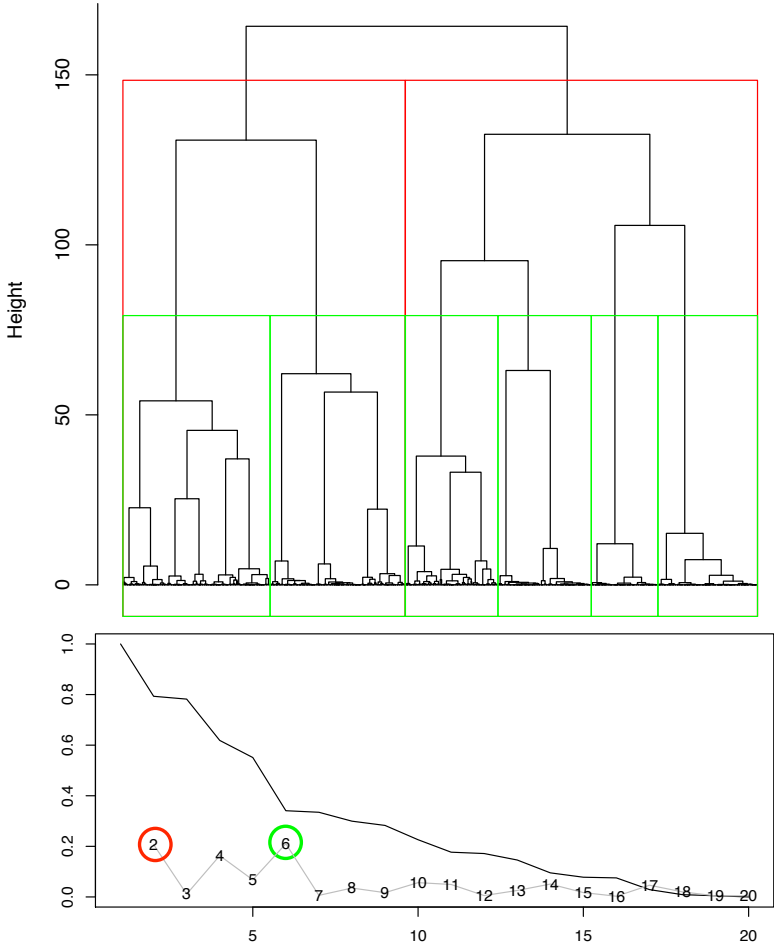
Figure 3. Motivations clusters: Box-plot for the six clusters solution.



Box plots in Figures 3 and 5 allow investigating on the distribution and interpretation of these clusters centres with respect to original segmentation variables. The red line that runs across all the box-plots of a cluster reports the sample mean of each variable. For the sake of interpretation, it is important to emphasize that the higher the height of the grey box (i.e. interquartile range), the smaller the homogeneity of the segment with respect to the variable

considered. This implies that segments are better characterised by those variables presenting low dispersion, and that the stronger the dispersions of variables among segments, the more dissimilar the segments.

Figure 4. Expenditures clusters: Bagged Clustering dendrogram together with the plot regarding the relative height of aggregation (black line) and the first differences (grey line).



4.1 Motivations clusters

Items concerning motivations asked how strongly each respondent agreed with a set of push factors for the visit to the CM. Each item used a 5-points Likert scale. The set included “shopping”, socialising with friends and relatives (“socialise”), fun/relax (“relax”), meeting new people (“meet”), doing something different and original (“do”), bringing my partner/family (“family”), supporting a local community initiative (“support”), tasting local products – food and beverages (“taste”), staying in a unique Christmas atmosphere (“atmosphere”), visitng the town centre in Merano (“town”), merry-go-round for children

(“merry-go-round”), train for children (“train”), pastry shop for children (“pastry shop”), ice-skating rink (“ice-skating”), band – musical group (“band”).

Figure 5. Expenditures clusters: Box-plot for the six clusters solution.

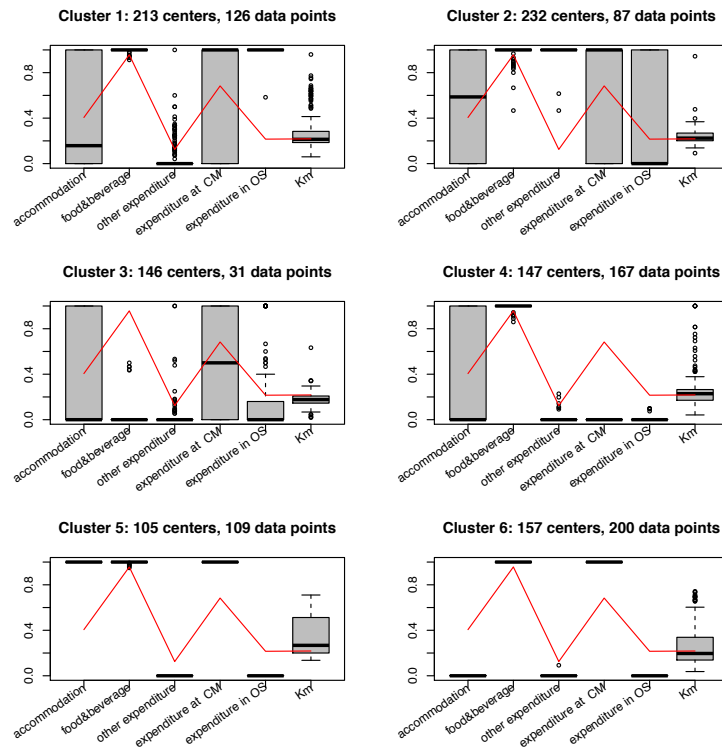


Table 2. Internal Validity index.

Partition	Motivations clusters			Expenditures clusters		
	Silhouette	Dunn	Davies-Bouldin	Silhouette	Dunn	Davies-Bouldin
2	0.412951	0.193641	1.167199	0.224247	0.232879	2.080234
3	0.237849	0.058901	1.900535	0.222397	0.232879	1.712915
4	0.221419	0.058901	1.597761	0.308009	0.230251	1.445356
5	0.195104	0.059349	1.685152	0.306704	0.230251	1.218014
6	0.208293	0.059349	1.557620	0.369999	0.266997	1.065399
7	0.208561	0.059349	1.652248	0.402177	0.266997	0.964006
8	0.203678	0.074163	1.716759	0.451496	0.266997	0.936226
9	0.220900	0.074163	1.656302	0.499022	0.271747	0.856294
10	0.218360	0.074163	1.599177	0.548231	0.197971	0.860568

Table 3 reports the median and average composition for each of the identified clusters. Kruskal–Wallis test was calculated in order to test the null hypothesis of equal median of each motivation variable among the six clusters. Together with Figure 3, it allows to investigate on the characteristics of the clusters.

Table 3. Motivations clusters: median value, average value in brackets, for the six clusters solution.

	Sample	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
shopping	3 (3.30)	3 (3.32)	3 (3.46)	4 (3.59)	3 (3.01)	4 (3.81)	2 (2.22)
socialise	3 (3.04)	5 (4.28)	1 (1.53)	3 (3.13)	3 (2.97)	4 (4.03)	4 (3.35)
relax	5 (4.24)	5 (4.53)	4 (3.92)	5 (4.40)	5 (4.27)	5 (4.62)	4 (3.63)
meet	3 (2.66)	4 (3.65)	1 (1.65)	3 (2.92)	3 (2.66)	4 (3.38)	1 (1.96)
do	5 (4.14)	5 (4.57)	4 (4.18)	5 (4.38)	5 (4.08)	5 (4.66)	3 (2.54)
family	4 (3.69)	5 (4.67)	3 (3.16)	5 (4.49)	5 (4.15)	1 (1.30)	2 (2.51)
support	3 (2.90)	4 (3.72)	1 (1.82)	4 (3.57)	3 (3.18)	1 (1.62)	3 (2.49)
taste	5 (4.57)	5 (4.67)	5 (4.47)	5 (4.68)	5 (4.55)	5 (4.75)	5 (4.25)
atmosphere	5 (4.62)	5 (4.75)	5 (4.81)	5 (4.68)	5 (4.66)	5 (4.86)	4 (3.70)
town	5 (4.44)	5 (4.75)	5 (4.51)	5 (4.39)	5 (4.54)	5 (4.52)	4 (3.98)
merry-go-round	1 (2.07)	5 (4.68)	1 (1.11)	1 (1.14)	4 (3.90)	1 (1.05)	1 (1.15)
train	1 (2.06)	5 (4.72)	1 (1.07)	1 (1.14)	4 (3.88)	1 (1.00)	1 (1.20)
pastry shop	1 (2.05)	5 (4.68)	1 (1.15)	1 (1.14)	4 (3.80)	1 (1.00)	1 (1.14)
ice-skating	1 (2.12)	5 (4.52)	1 (1.47)	1 (1.85)	2 (2.41)	1 (1.84)	1 (1.35)
band	3 (2.55)	4 (4.22)	1 (1.87)	2 (2.41)	3 (2.73)	3 (2.62)	1 (1.99)

Notes: Kruskal–Wallis test with ties was significant at $p \leq 0.01$.

Tasting local foods and beverages (“taste”) is the most important motivation in visiting the Merano CM, with no large differences among visitors. Other very important factors for five out of six groups are staying in a unique Christmas atmosphere (“atmosphere”), visiting the town centre of Merano (“town”), doing something different and original (“do”). However, visitors of cluster 6 paid less attention to these four aspects. The overall score assigned to the children’s attractions (“merry-go-round”, “train”, “pastry shop”) allows to distinguish the clusters into two types: that in which the travel group included children (cluster 1, and 4 to a lesser extent), and that where the children’s attractions were more or less absent (the remainder). As expected, visitors of clusters 1 and 4 considered staying with their partner/family as important when they visit this type of cultural event. Also members of cluster 3 gave importance to family, but they probably did the visit only with their partner, or they did not come for the children attractions. Shopping is not a fundamental motivation since no cluster reports a high value. However cluster 5 and 6 show respectively the highest and lowest average scores about its importance for attending CM.

As regards to the other attractions proposed by the organization of the CM (“ice-skating” and “band”), only the visitors of cluster 1 declared themselves to be interested. Socialising with friends and relatives (“socialise”) is quite important for visitors of clusters 1 and 5, whereas it has a low importance to respondents classified in group 2. No cluster shows particular homogeneity in attributing importance to shopping.

Summarizing the results, cluster 1 (“active families”) is likely made by families with children who are actively involved in the atmosphere and activities of the market. Cluster 2 (“CM and place seekers”) did not come to socialize with and meet new people, support local community, and attending Christmas attractions, but their highest motivation is the atmosphere of an event they think it is different and original, as well as visiting the town. Cluster 3 (“families not interested in children’s attractions”) is characterised by families/couples without children, or whose interest does not rely on children’s attractions. Cluster 4 (“families seeking CM and place”) is similar to cluster 3, but they did not declare a negative attitude towards children’s attractions. Cluster 5 (“singles seeking CM and place”) is similar to cluster 2, and in addition it is composed by people who did not come to the market to bring a member of the family, and do not necessarily give the same low importance to CM attractions. All these five clusters emphasize uniqueness in the atmosphere of the event. Indeed, the popularity of CMs is based on the idea of a place that is unique in a specific time of the year and place (i.e., the ones with German culture), where people can consume local products, some of which are traditionally related to Christmas time. Consistently, the five Clusters provide different declinations of a positive attitude towards the market’s atmosphere, where Cluster 1 shows the highest number of items with the highest and less variable score.

On the opposite side, cluster 6 aggregates visitors who paid less attention to shopping than the others, doing other particular activities during the visit, and staying in a unique Christmas atmosphere, though to a certain extent they reveal attention to local food and beverages. This last segment clusters people whose attitude contrasts with the common idea of the CM as a place for shopping in a unique atmosphere, unlike all the other five groups. Consistently with the aim of the paper, further analyses will be proposed in order to investigate this cluster more in depth. Better understanding of such “extreme” segment, i.e. the one whose motivations contrast with the others, can provide useful policy indications in order to find reasons to improve visibility and impact even to less motivated people.

4.2 Expenditures clusters

Six expenditure categories were taken into consideration in order to segment visitors. Five of them related to the declared level of spending of each visitor for a particular item: accommodation per night per person (“accommodation”), food and beverage per day per person (“food & beverage”), other expenditure (pharmacy/health, tour guide service, local transportation, other expenditure in Merano); shopping at the CM (“expenditure at CM”); shopping in other places of the city (“expenditure in OS”). A sixth category included the

distance in kilometres (“km”) between the city of residence and Merano as proxy for transportation costs. For segmentation purposes, each quantitative variable is transformed into a dummy variable, equalling 1 when a positive expenditure is observed and 0 otherwise. This transformation was done in order to group visitors who spent at least some money on each of the recorded items, and therefore had positive attitude in spending, rather than putting together those who spent similar amounts of money. This perspective can be of greater interest in order to describe the behaviour of those who did not record any amount of money in some categories.

For the sake of comparison, the transportation costs were normalized to the [0,1] interval, by dividing this variable by the maximum value observed in its empirical distribution. The original information on levels of spending concurred in forming average expenditure that were tested through ANOVA tests, which aimed at verifying whether expenditure variables reported equal means among the six clusters observed (Table 4). Also here, results should be read jointly with what reported by Figure 5.

Table 4. Expenditure clusters: average expenditure by category of expenditure in €, and percentage of positive expenditure in parenthesis for the six clusters solution.

Expenditure	Sample	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
km	339.73	350.97	372.76	285.48	312.26	404.44	300.46
accommodation	18.96 (40.55)	19.15 (42.06)	26.8 (59.77)	14.38 (22.58)	17.62 (42.51)	50.75 (100)	0.00 (0)
food & beverage	23.22 (95.83)	25.94 (100)	25.18 (100)	0.00 (0)	25.17 (100)	25.06 (100)	22.31 (100)
other	1.74 (12.27)	0.08 (1.59)	14.56 (100)	0.16 (3.23)	0.00 (0)	0.00 (0)	0.00 (0)
shop at the CM	40.76 (68.45)	61.37 (84.92)	39.83 (71.26)	30.97 (45.16)	0.00 (0)	72.24 (100)	49.45 (100)
shop in OS of the city	17.52 (21.22)	78.4 (100)	20.52 (29.89)	6.77 (9.68)	0.00 (0)	0.00 (0)	0.00 (0)

Notes: ANOVA test was significant at $p \leq 0.01$.

All visitors in cluster 1 (named “Pro shopping”) spent money in food and beverage, other shops of the city and recorded the highest average amount for this latter category. In addition they show the second highest average expenditure level at the shop of CM. Interviewees of cluster 2 (called “Spenders”) spent in food and beverage, report the highest expenditure level in “other” items, and spent a non-indifferent amount of money also in the remaining categories. Cluster 3 (“Low consume”) is the smallest (4.3% of the entire sample) and it groups visitors who, unlike the other visitors, did not spend on “food & beverage” and other

items, spent significantly less than the other visitors in all the remaining expenditure categories, and came mainly from places near Merano. Cluster 4 (“No shopping”) aggregates visitors who spent money on “food & beverage”, some money on “accommodation”, and avoided shopping at the CM as well as and expenditure on other categories. All visitors grouped in cluster 5 (“Tourists pro shop CM”) and 6 (“Day-visitors pro shop CM”) had a positive expenditure at the CM with an average expenditure respectively equal to €72.24 and €49.45. Differences arise because visitors in cluster 5 spent on average the highest amount for accommodation (€50.75), stayed at least one night outside home, and came from the furthest destinations. Interviewees in cluster 6 were instead day-visitors since they did not spend money on accommodation.

Further analyses will be proposed in order to investigate these last two clusters more in depth. Together with transportation, accommodation constitutes the item where tourists spend the highest amount of money. Unlike transportation, accommodation facilities have a high and significant impact on local communities, in terms of occupation and spin-off effects to satellite activities. To this end, two opposite situations were analysed, i.e. the cluster of tourists spending the highest amount on accommodation, and same day visitors.

Table 5. Comparison between motivations and expenditures clusters.

		Motivations clusters						Total
		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	
Expenditures clusters	Cluster1	14 (21.54)	24 (17.78)	40 (17.78)	24 (16.22)	11 (15.71)	11 (15.07)	124 (17.32)
	Cluster2	7 (10.77)	17 (12.59)	27 (12)	16 (10.81)	7 (10)	12 (16.44)	86 (12.01)
	Cluster3	5 (7.69)	5 (3.70)	7 (3.11)	8 (5.41)	1 (1.43)	5 (6.85)	31 (4.33)
	Cluster4	13 (20)	36 (26.67)	57 (25.33)	31 (20.95)	13 (18.57)	17 (23.29)	167 (23.32)
	Cluster5	13 (20)	18 (13.33)	34 (15.11)	29 (19.59)	9 (12.86)	6 (8.22)	109 (15.22)
	Cluster6	13 (20)	35 (25.93)	60 (26.67)	40 (27.03)	29 (41.43)	22 (30.14)	199 (27.79)
	Total	65 (100)	135 (100)	225 (100)	148 (100)	70 (100)	73 (100)	716 (100)

Finally, interesting evidence derives from Table 5, where the two classifications are compared through a cross-tabulation and Fisher’s exact test. Results indicate the independence of the classifications adopted (p -value=0.579). This suggests that similar expenditure profiles in terms of per capita levels may not be associated with the motivations profiles.

4.3 Clusters description

As reported previously, the composition of specific behaviours was investigated further. From the motivations side, deeper investigation concerned the segment named “No CM” (cluster 6). Regarding expenditures clusters two segments were considered: “Tourists pro

shop CM” (cluster 5), who group visitors with the highest and less variable level of spending in accommodation and the highest one of the expenditure at the CM; “Day visitors pro shop CM” (Cluster 6) a very particular segment of visitors of events who did not spend money on accommodation though they made purchases at the CM.

Membership to the clusters was modelled via two different procedures. Binary logistic regression was used to test the significance of a set of variables on being part of the motivations cluster “No CM”. Multinomial logistic regression analysis was instead conducted to determine if “Tourists pro shop CM” and “Day-visitors pro shop CM” are simultaneously different in terms of personal characteristics from other visitors. A set of socio-demographic and economic variables (gender, age, level of education, occupation, household status, income, distance covered), and measures of travel-related behaviours (days of the week in which the visit took place, number of visits to the CM, number of nights outside home, visiting party composition) were used as regressors (the description of these independent variables is given in the Appendix). Resulting coefficients are reported in Table 6.

Different variables discriminated significantly between being member of the “No CM” cluster and not. The higher the age of the visitors (“Age”), the lower (less than proportional) the probability of being members of “No CM” cluster. This indicates that elder people are more interested in the Christmas atmosphere and in shopping than the younger. Visitors with a “Permanent contract (part-time)” and those who did not state their family income level (“Missing income”) are significantly more likely to be a member of “No CM” cluster. Those who visited the CM with an “Organised group” were less likely to be disinterested in the Christmas atmosphere and in shopping, perhaps because they paid a fee for the organized trip, did shopping, and “brought home a bit of Christmas” through one of the souvenirs of the shop. Finally, the negative relationship between the distance (“Km”) from the place of residence and “non interest” in CM can be easy to explain since visitors coming from far places made a long trip to visit something they considered worthwhile.

Concerning “expenditures clusters”, only trip related variables resulted in significantly affecting the membership to the two segments, whereas socio-demographic and economic characteristics did not appear as significant. As expected, length of stay (“Number of nights”) affected the likelihood of being a member of the two clusters, even if in a different way – i.e., tourists tend to stay longer, whereas number of days is negatively related to same-day visiting. Members of cluster 5 who already visited the CM in Merano are less likely to return and stay overnight. This can be mainly due to the similarity of “products of the CM” from one year to another, as well as of the “CM as a product”. The same happens to day-visitors, though to a

lesser extent. The likelihood of being an overnight stayer also increases if the interview was conducted during the “Week-end” (Friday, Saturday, or Sunday), because the probability of staying on holiday for a longer period is higher if the week-end is included. We also observe a positive influence of the weekend on same day-visiting. Finally, those who visited the CM with an organised group were more likely to be same day-visitors, perhaps because a one-day journey to a place with such a kind of attraction is more pleasurable if shared with other people.

Table 6. Regressions results.

Independent variables	Motivations clusters ^A		Expenditures clusters ^B	
	Cluster 6		Cluster 5	Cluster 6
Week-end	-0.346 (0.37)		0.903 (0.39)**	0.136 (0.31)
Number of past visits to Merano CM	-0.033 (0.05)		-0.215 (0.10)**	-0.016 (0.06)
Number of nights	-0.142 (0.10)		0.295 (0.17)*	-1.198 (0.20)***
Distance in Km ^C	-0.003 (>0.01)**		–	–
Visiting party				
Couple	0.072 (0.34)		0.485 (0.34)	-0.018 (0.26)
Children	-0.167 (0.45)		-0.158 (0.43)	-0.126 (0.34)
Organised group	-1.605 (0.70)**		0.709 (0.44)	0.817 (0.31)***
Number of family trip	-0.141 (0.16)		0.054 (0.09)	-0.047 (0.08)
Household situation				
Female	0.320 (0.30)		-0.211 (0.24)	0.026 (0.22)
Age	-0.192 (0.09)**		0.148 (0.11)	0.028 (0.07)
Age ²	0.003 (>0.00)***		-0.002 (>0.00)	0.001 (>0.00)
High school	0.323 (0.45)		0.161 (0.37)	0.117 (0.34)
University degree or more	0.421 (0.48)		-0.026 (0.41)	-0.059 (0.38)
Single	-0.497 (0.51)		-0.153 (0.74)	-0.600 (0.48)
Married	-0.641 (0.47)		0.055 (0.69)	-0.694 (0.44)
Autonomous worker	0.617 (0.56)		0.170 (0.51)	0.195 (0.41)
Permanent contract (full time)	0.628 (0.48)		0.134 (0.41)	0.406 (0.35)
Permanent contract (part time)	1.260 (0.56)**		-0.551 (0.58)	0.085 (0.46)
Fixed contract	0.607 (0.56)		0.023 (0.56)	0.584 (0.41)
Income ^D	–		0.001 (>0.00)	0.001 (>0.00)
Missing income	0.675 (0.28)**		–	–
Constant	1.643 (1.93)		-5.958 (2.42)**	-0.793 (1.53)

Notes:

Robust Standard Error in parenthesis. * p<0.10; ** p<0.05; *** p<0.01;

^A Logistic regression: N = 745; Wald chi2(20) = 54.53; Prob > chi2 = 0.00; Pseudo R²=0.1212; McFadden R² = 0.121; Cox & Snell R² = 0.076; Nagelkerke R² = 0.159; McKelvey & Zavoina's R² = 0.268.

^B Multinomial logit: N = 715; Wald chi2(38) = 95.78; Prob > chi2 = 0.00; Pseudo R²=0.1876; McFadden R² = 0.188; Cox & Snell R² = 0.303; Nagelkerke R² = 0.355.

^C Not inserted in expenditure clusters regression because it concurred in forming clusters.

^D Not inserted in motivations clusters regression because all visitors in this cluster reported the same income.

^E Not inserted in expenditure clusters regression because all visitors in these clusters declared their income.

5. Conclusions

This paper is the first attempt to apply BC to the analysis of cultural events. Although this technique is not exempt from limitations, it provides several advantages with respect to other

classic unsupervised methods. BC on a survey of tourists at the 2011 CM of Merano, a city in the Northern Italy, pointed out different interesting features of the behaviour of visitors. Unlike other previous papers clustering attendees of cultural events, this work adopted of a twofold perspective. CMs are events that attract visitors in search of a “typical” atmosphere, which also have an important economic impact on local communities. For these reasons, clusters were formed from motivations and expenditure items. A first interesting evidence concerned the absence of association between the two clusters: motivations segments are unlikely to correspond to expenditure behaviours. This suggests to treat the two classifications separately. However, some elements in common arise.

Results about motivations sketched a more detailed behaviour of each segment than what emerged by the only similar paper on CM by Brida et al. (2013a). The latter used Self-Organised Maps (SOM) to find three segments of business people, Christmas fans, and general tourists. The latter two groups aggregate visitors who declared positive interest towards the CM. Similarly, five out of six clusters of the present paper reported positive and very high scores for aspects that are strictly related to CM, as uniqueness of the atmosphere and taste of typical food and beverage. The main caveat is that Brida et al. (2013a) employed different sets of variables than ours. In fact, the present paper included only motivation-related variables, thus excluding motivations related to the purpose of the visit to the city such as business. However, parallelism emerges between their “business people” segment and our “no CM” one, in terms of low interest towards this cultural event. Indeed, the importance of the business segment as a distinct one is expected because two out of the three cities of Brida et al.’s (2013a) survey were the main ones of the region (i.e., Trento and Bolzano). This paper instead intended to reduce the limit of the analysis to one single event.

In general our empirical evidence is consistent with a part of the literature stressing family ties as main elements identifying the segments (among others, Lee et al, 2004; Li et al., 2009). Clusters can be distinguished according to the group composition and the presence of motivations related to family and children. However, general agreement on the importance of some factors that are typical of the common way to mean a CM emerged in five out of six clusters. For this reason, the presence of a group that is poorly related to all these factors deserved attention. Indeed, the situation of having most of the interviewed people with motivations that are consistent with the “spirit” of the CM helps policymakers to address *ad-hoc* policies to attract them further. The investigation on the cluster of “no CM” people can help to develop and expand the interest on event towards people with different motivations. This group is essentially made of young people who are not organized in group and leave

nearby. To increase the interest of this peculiar segment as well as their economic impact, other collateral events may be organized and opportunely advertised. This is the direction that is suggested also by the low score on push factors on attractions like ice-skating and band, which is reported by “CM and place seekers” besides “no CM”, as well as the relatively low score of other segments.

As to expenditure clusters, we investigated those which experimented two opposite behaviours in terms of accommodation expenditure, that is overnight stayers and same day visitors. Both groups spend in shopping. However the overnight stayers cluster, named “Tourists pro shop CM”, is made of people who are likely to spend the highest amount at the shops. One of the most interesting features is that these people unlikely visited the market for a high number of times. Indeed, stimulating the repeat visit is a main goal also of cultural attractions, also because of the implications it has on local economies (Brida et al., 2013c). This recalls what was stressed above, about the need to organize a set of activities and collateral events in order to promote the novelty in revisiting CM. Therefore, innovating a traditional attraction, with events that can be attractive for all segments, is a policy that seems to be suggested also by the two expenditure clusters we analysed. In fact, the absence of a form of loyalty occurs also in the segment of same day visitors who leave in near locations.

Of course, if the promotion of new events and attractions from one side can increase the variety of people attracted by the event, on the other side the effect of these policies towards the customers that care about a more traditional way to mean CM has to be verified. Our results seem also to propose other directions for the aim of promoting and enhance the economic impact of the event. Reading the results of logistic regression in cluster “no CM”, it can be suggested that non-young visitors from far places can be a segment of people which is not indifferent to what a CM is aimed to propose as a cultural event. Searching for something that is proper of other cultures is a motive for many visitors to come from far places. If the risk of introducing non-traditional events and attractions can even question the idea of the CM as cultural event, a proper promotion on far regions with different culture can increase the flow of those in search of a typical cultural event.

In addition the highest economic impact to the CM is significantly related to the presence of overnight stayers during week-ends. Also, same-day visitors spending in shopping at CM visit it in this part of the week. Certainly, the weekends of a CM are the times when markets are overcrowded. Working on reducing the seasonality, and have a less homogeneous distribution of the number of people during the week, means to provide incentives to that shares of the population who might be available to visit it also in other days.

This study presents some limitations. The use of non-probabilistic samples might have affected the results. Also regarding the data collection method, surveying people at a CM has the advantage of profiling people who are most typically interested in it. Indeed, a representative sample would have given further information about the significance of the obtained segments. Inferring the opinions of a whole population about the interest to the CM could have provided more robust marketing indications. In this sense, the application of a representative design is a research direction of interest. In addition, the same survey methods of this paper can be applied to other similar contexts for the sake of testing whether such results are stable over space and time.

Acknowledgments

This research was supported by the Autonomous Province of Bolzano project “Le attrazioni culturali e naturali come motore dello sviluppo turistico. Un’analisi del loro impatto economico, sociale e culturale”, Research Funds 2009; and by the Free University of Bolzano projects “The museums’ visitors: a comparative socio-economic analysis” and “Determinants of tourist expenditure: theory and microeconomic models”. Authors would like to thank the participants to the XLIX Annual Conference of the Italian Society of Economics, Demographics, and Statistics (SIEDS) for their very valuable comments.

Appendix

Description of the explanatory variables.

Independent variables	Descriptions	Mean value
Socio-demographic and economic characteristics		
Female	1= female; 0= male	0.58
Age	Age of the respondent (continuous)	38.27
Age^2	Squared age of the respondent (continuous)	1569.58
High school	1 = education level is high school or postgraduate; 0 = otherwise	0.57
University	1 = education level is university degree or postgraduate; 0 = otherwise	0.30
Single	1 = Single; 0 = otherwise	0.29
Married	1 = Married; 0 = otherwise	0.66
Autonomous worker	1= autonomous worker and employed (full-time or part-time); 0 = otherwise	0.14
Permanent contract (full time)	1= employed with a permanent contract full-time; 0 = otherwise	0.52
Permanent contract (part time)	1= employed with a permanent contract part-time; 0 = otherwise	0.09
Fixed contract	1= employed with a fixed contract; 0 = otherwise	0.11
Distance in Km	distance in kilometres between the city of residence and the city that hosts the CM (continuous)	339.73
Income	The central values of the following income categories in € were considered: 0 - 25,000; 25,000 - 50,000; 50,000 - 75,000; 75,000 – 100,000; 100,000 – 150,000; > 150,000; 0 if the respondent does not declare her income (continuous)	26122.19
Missing income	1 = respondent does not declare his/her income; 0 = otherwise	0.26
Characteristics of the visit		
Week-end	1= the interview was made on Friday, Saturday or Sunday; 0= otherwise	0.86
Number of visit to the CM in Merano	Number of CM in Merano visited in the past (discrete)	0.92
Number of nights	Length of stay in Merano (discrete)	1.38
Visiting party		
Couple	1 = visited with the partner/spouse; 0 = otherwise	0.44
Children	1 = visited with children; 0 = otherwise	0.36
Organised group	1 = the visit at the CM was made with an organized group and/or with friends/colleagues; 0 = otherwise	0.12
Number of family trip	Number of members of the family doing the visit of the CM with the interviewee	1.54

Notes: In the case of dichotomous variables, the mean value is equal to the proportion of 1.

References

- Albalade, D., & Bel, G. (2010). Tourism and urban public transport: Holding demand pressure under supply constraints. *Tourism Management*, *31*, 425–433.
- Beane, T. T., & Ennis, D. M. (1987). Market segmentation: A review. *European Journal of Marketing*, *21*, 20–42.
- Bigné, J. E., & Andreu, L. (2004). Amotions in Segmentation. An Empirical Study. *Annals of Tourism Research*, *31*, 682–696.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*, 123–140.
- Brida, J. G., Disegna, M., & Osti, L. (2012a). Authenticity Perception of Cultural Events: A Host-Tourist Analysis. *Tourism, Culture and Communication*, *12*, 85–96.
- Brida, J. G., Disegna, M., & Osti, L. (2012b). Segmenting visitors of cultural events by motivation: A sequential non-linear clustering analysis of Italian Christmas Market visitors. *Expert Systems with Application*, *39*, 11349–11356.
- Brida, J. G., Disegna, M., & Osti, L. (2013a). The Effect of Authenticity on Tourists' Expenditure at Cultural Events. *Current Issues in Tourism*, *16*, 266–285.
- Brida, J. G., Disegna, M., & Scuderi, R. (2013b). Visitors of two types of museums: a segmentation study. *Expert Systems with Applications*, *40*, 2224–2232.
- Brida, J. G., Disegna, M., & Scuderi, R. (2013c). The behaviour of repeat visitors to museums: review and empirical findings. *Quality & Quantity*, doi: 10.1007/s11135-013-9927-0.
- Chang, J. (2006). Segmenting tourists to aboriginal cultural festivals: An example in the Rukai tribal area, Taiwan. *Tourism Management*, *27*, 1224–1234.
- Chen, J. S., & Hsu, C. H. C. (1999). The use of logit analysis to enhance market segmentation methodology. *Journal of Hospitality & Tourism Research*, *23*, 268–283.
- Choi, A. S. (2011). Implicit prices for longer temporary exhibitions in a heritage site and a test of preference heterogeneity: A segmentation-based approach. *Tourism Management*, *32*, 511–519.
- Claver-Cortés, E., Molina-Azorín, J. F., & Pereira-Moliner, J. (2007). Competitiveness in mass tourism. *Annals of Tourism Research*, *34*, 727–745.
- Cochran, W. G. (1977). *Sampling techniques*. Third edition, Wiley series in probability and mathematical statistics, John Wiley & Sons

- Cohen, S. A., Prayag, G., & Moital, M. (2014). Consumer behaviour in tourism: Concepts, influences and opportunities. *Current Issues in Tourism*. DOI:10.1080/13683500.2013.850064
- Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Trans. Pattern Anal. Machine Intell.*, 1, 224–227.
- De Guzman, A. B., Leones, J. D., Tapia, K. K. L., Wong, W. G., & de Castro, B. V. (2006). Segmenting motivation. *Annals of Tourism Research*, 33, 863–867.
- Dolnicar, S. (2002). A review of data-driven market segmentation in tourism. *Journal of Travel & Tourism Marketing*, 12, 1–22.
- Dolnicar, S. (2003). Using cluster analysis for market segmentation—typical misconceptions, established methodological weaknesses and some recommendations for improvement. *Australasian Journal of Market Research*, 11, 5–12.
- Dolnicar, S., & Leisch, F. (2000). Getting More Out of Binary Data: Segmenting Markets by Bagged Clustering. *Working Paper Series 71 SFB “Adaptive Information Systems and Modeling in Economics and Management Science”*, <http://www.wu-wien.ac.at/am>, August 2000.
- Dolnicar, S., & Leisch, F. (2003). Winter Tourist Segments in Austria: Identifying Stable Vacation Styles Using Bagged Clustering Techniques. *Journal of Travel Research*, 41, 281–292.
- Dolnicar, S., & Leisch, F. (2004). Segmenting Markets by Bagged Clustering. *Australasian Marketing Journal*, 12, 51–65.
- Dolnicar, S., Crouch, G. I., Devinney, T., Huybers, T., Louviere, J. J., & Oppewal, H. (2008). Tourism and discretionary income allocation. Heterogeneity among households. *Tourism Management*, 29, 44–52.
- Dunn, J. C. (1974). Well-separated clusters and fuzzy partitions. *Journal of Cybernetics*, 4, 95–104.
- Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis*. Wiley series in Probability and Statistics, London.
- Franke, N., Reisinger, H., & Hoppe, D. (2009). Remaining within-cluster heterogeneity: a meta-analysis of the “dark-side” of clustering methods. *Journal of Marketing Management*, 25, 273–293.
- Handl, J., Knowles, J., & Kell, D. B. (2005). Computational Cluster Validation in Post-Genomic Data Analysis. *Bioinformatics*, 21, 3201–3212.

- Hughes, H. L. (2002). Culture and tourism: A framework for further analysis. *Managing Leisure*, 7, 164–175.
- Jain, A. K. (2010). Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 31, 651–666.
- Kang, K., Hua-Xiang, Z., & Ying, F. (2008). A novel Cluster Ensemble Algorithm Based on Dynamic Cooperation. *Fifth International Conference on Fuzzy Systems and Knowledge Discovery, IEEE Computer Society*, 32–35.
- Kau, A. K., & Lim, P. S. (2005). Clustering of Chinese tourists to Singapore: an analysis of their motivations, values and satisfaction. *Journal of Vacation Marketing*, 7, 231–248.
- Koc, E., & Altinay, G. (2007). An analysis of seasonality in monthly per person tourist spending in Turkish inbound tourism from a market segmentation perspective. *Tourism Management*, 28, 227–237.
- Konu, H., Laukkanen, T., & Komppula, R. (2011). Using ski destination choice criteria to segment Finnish ski resort customers. *Tourism Management*, 32, 1096–1105.
- Kotler, P., & Armstrong, G. (1999). *Principles of Marketing* 8th ed. Prentice Hall.
- Kotler, P., Bowen, J. T., & Makens, J. C. (2010). *Marketing for hospitality and tourism* (5th Edn.). Upper Saddle River, New Jersey: Pearson Prentice Hall.
- Kruger, M., Saayman, M., & Ellis, S. (2011). Segmentation by genres: the case of the Aardklop national arts festival. *International Journal of Tourism Research*, 13, 511–526.
- Lee, C. K., Lee, Y. K., & Wicks, B. E. (2004). Segmentation of festival motivation by nationality and satisfaction. *Tourism Management*, 25, 61–70.
- Lee, J., & Beeler, C. (2009). An investigation of predictors of satisfaction and future intention: links to motivation, involvement, and service quality in a local festival. *Event Management*, 13, 17–31.
- Lee, C. K., & Lee, T. H. (2001). World Culture EXPO segment characteristics. *Annals of Tourism Research*, 28, 812–816.
- Leisch, F. (1999). Bagged Clustering. *Working paper Series 51 SFB “Adaptive Information Systems and Modeling in Economics and Management Science”*, <http://www.wu-wien.ac.at/am>, August 1999.
- Leisch, F. (2006). A toolbox for K-centroids cluster analysis. *Computational Statistics & Data Analysis*, 51, 526–544.
- Li, M., Huang, Z., & Cai, L. A. (2009). Benefit segmentation of visitors to a rural community-based festival. *Journal of Travel & Tourism Marketing*, 26, 585–598.

- Liao, S., Chu, P., & Hsiao, P. (2012). Data mining techniques and applications – A decade review from 2000 to 2011. *Expert Systems with Applications*, 36, 11772–11781.
- Najmi, M., Sharbatoghlie, A., & Jafarieh, A. (2010). Tourism Market Segmentation in Iran. *International Journal of Tourism Research*, 12, 497–509.
- Nicholson, R., & Pearce, D. (2000). Who goes to events: A comparative analysis of the profile characteristics of visitors to four South Island events in New Zealand. *Journal of Vacation Marketing*, 6, 236–253.
- Oom do Valle, P. O., Mendes, J., & Guerreiro, M. (2010). Sustainable cultural events based on marketing segmentation: The case of Faro capital of culture. *Pasos*, 8, 91–104.
- Pérez, E. A., & Nadal, J. R. (2005). Host Community Perceptions. A Cluster Analysis. *Annals of Tourism Research*, 32, 925–941.
- Pesonen, J. A. (2013). Information and communications technology and market segmentation in tourism: a review. *Tourism Review*, 68, 14–30.
- Prentice, R., & Andersen, V. (2003). Festival as creative destination. *Annals of Tourism Research*, 30, 7–30.
- Punj, G., & Steward, D. W. (1983). Cluster analysis in marketing research: Review and suggestions for application. *Journal of Marketing Research*, 20, 138–148.
- Rousseeuw, P. J., (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65.
- Thompson, K., & Schofield, P. (2009). Segmenting and profiling visitors to the Ulaanbaatar Nadaam festival by motivation. *Event Management*, 13, 1–15.
- Tkaczynski, A., & Rudle-Thiele, S. R. (2011). Event segmentation: a review and research agenda. *Tourism Management*, 32, 426–434.
- Tuma, M. N., Decker, R., & Scholz, S. W. (2011). A survey of the challenges and pitfalls of cluster analysis application in market segmentation. *International Journal of Market Research*, 53, 391–414.
- Vesanto, J., & Alhoniemi, E. (2000). Clustering of the self-organizing map. *IEEE Transactions on Neural Networks*, 11, 586–600.
- Wedel, M., & Steenkamp, J. (1989). A fuzzy clusterwise regression approach to benefit segmentation. *International Journal of Research in Marketing*, 6, 241–258.