

## VISITORS OF TWO TYPES OF MUSEUMS: A SEGMENTATION STUDY

Juan Gabriel Brida; Marta Disegna<sup>\*</sup>; Raffaele Scuderi

School of Economics and Management

JuanGabriel.Brida@unibz.it; marta.disegna@unibz.it; Raffaele.Scuderi@unibz.it

Free University of Bozen-Bolzano

Universitätsplatz 1 - piazza Università, 1

39100 Bozen-Bolzano – Italy

### **Abstract**

Market segmentation comprises a wide range of measurement tools that are useful for the sake of supporting marketing and promotional policies also in the sector of cultural economics. This paper aims to contribute to the literature on segmenting cultural visitors by using the Bagged Clustering method, as an alternative and effective strategy to conduct cluster analysis when binary variables are used. The technique is a combination of hierarchical and partitioning methods and presents several advantages with respect to more standard techniques, such as k-means and LVQ. For this purpose, two ad-hoc surveys were conducted between June and September 2011 in the two principal museums of the two provinces of the Trentino-South Tyrol region (Bolzano and Trento), Northern Italy: the South Tyrol Museum of Archaeology in Bolzano (ÖTZI), hosting the permanent exhibition of the “Iceman” Ötzi, and the Museum of Modern and Contemporaneous Art of Trento and Rovereto (MART). The segmentation analysis was conducted separately for the two kinds of museums in order to find similarities and differences in behaviour patterns and characteristics of visitors. The analysis identified three and two cluster segments respectively for the MART and ÖTZI visitors, where two ÖTZI clusters presented similar characteristics to two out of three MART groups. Conclusions highlight marketing and managerial implications for a better direction of the museums.

*Keywords:* Bagged clustering, Logit models, Museum, Segmentation, Motivation.

*\*Corresponding Author*

## 1. Introduction

Museums are the most popular cultural attractions, usually followed by art galleries and monuments (McKercher, 2004). For a long time visitors of cultural attractions were treated as a homogeneous mass of people. The tendency of the recent tourism literature is instead to consider them as a heterogeneous market with different characteristics, perceptions and needs (Hughes, 2002). Brida et al. (2012) showed that visitors of Christmas Markets in Northern Italy clustered into three groups according to a set of motivational factors that drove them to make the visit. Other studies showed that tourists who visited art museums presented different socio-demographic characteristics (in particular regarding the level of education, income and occupation) than those who engaged in festivals, musical activities, theme parks, amusement parks, local fairs, and events (Kim et al., 2007; Bennett, 1994; Schuster, 1991).

Most research on tourism considered different types of museums (like art museum, stamps, history, science, and even children's museums) as a unique cultural attraction with the same "label". However, MacDonald and Alford (1995) suggested they are heterogeneous, by affirming that "all museums are products of their particular cultural and historical experiences". Each museum exhibits its peculiarity by offering visitors different kinds of involvements (Dicks, 2003) and experiences, which are suitable for different kinds of tourists. Furthermore, an art museum, a history museum, an opera, or an outdoor festival might produce different experiences in visitors (Stylianou-Lambert, 2011). For these reasons research should analyse cultural attractions, and in particular museums, separately according to the subject matter and the experiences that they offer (Stylianou-Lambert, 2011).

Profiling museum visitors by taking into consideration also the different characteristics of the museums can be of crucial importance for managers and marketing analysts of the museum. Identifying homogeneous clusters of consumers-visitors can be in fact an essential step for planning and developing appropriate strategies, in order to satisfy the needs of each segment of guests. In this context clustering proposes a set of widely used unsupervised techniques with the aim to discover hidden associations among statistical units and identifying segments (Saarenvirta, 1998). Given a set of selected segmentation variables, these methodologies aggregate the units in groups, in such a way that each aggregation contains the most similar units, and at the same time is dissimilar from the remainder. The supervision means that "membership of data points which can illustrate the general structure of the group is required in order to derive the classification rules", therefore the supervision implies that there is no rule for initiation of classification (Budayan, Dikmen & Birgonul,

2009). This implies that the empirical distribution and characteristics of the data will determine the cluster membership.

Since the introduction of market segmentation in the late 1950s, the number and type of approaches for segmentation has grown enormously (Liao, Chu, & Hsiao, 2012; Dolnicar & Leisch, 2004). Unfortunately, as emphasized by many researchers, no absolutely “correct” way to segment a market exists in the literature (Brida, Disegna, & Osti, 2012; Kotler, Bowen, & Makens, 2010; Dolnicar et al., 2008; Beane & Ennis, 1987). On the contrary, the researcher intervenes in different moments of the estimation process, which of course involves the final results. This implies that “clustering is exploratory data analysis and different methods present different views of data” (Leisch, 2006). The degrees of freedom in the clustering algorithm concern, among other things, the variables selection, the choice of a measure of dissimilarity between units, the final number of clusters, the test of the clustering solution as not purely random, the interpretation of final results for addressing management and marketing. Moreover, one has to bear in mind that “in the case of no clear cluster structure there is no “correct” solution” (Leisch, 2006).

The most popular clustering techniques are partitioning and hierarchical methods. The standard partitioning procedures aim to group the observations around a centre in order to find a segmentation of a set of units in an *a priori* fixed number of clusters. In the marketing and tourism literature *k*-means is the most commonly used algorithm that falls into this category. Hierarchical methods instead obtain the final clusters solution by repeatedly joining the “closest” clusters composed of one or more observations (agglomerative clustering), or repeatedly splitting the “further” clusters (divisive clustering). This study instead makes use of Bagged Clustering, which combines both hierarchical and partitioning methods. It was proposed by Leisch (1999) and has the advantages of overcoming many of the limitations of the two methods. This method has been used successfully in the past by Leisch himself or his research team, for the sake of tourism market segmentation (Dolnicar & Leisch, 2000, 2003; Dolnicar et al., 2008), but it has been applied infrequently by other researchers in the same field or in others (Huang, Chang, & Wu, 2009). Its application to the field of culture aims to study the profiles of tourists with respect to their motivations in visiting two different types of museums. This can shed light on investigating whether museums offering different experiences are visited by heterogeneous types of tourists, or on the contrary if segments with common characteristics can be detected. This objective is pursued by using a dataset from ad-hoc surveys. These were conducted from June to September 2011 in the two main museums of Trento and Bolzano, the two provinces of the Trentino-South Tyrol region. The South

Tyrol Museum of Archaeology (shortened to ÖTZI) is located in the Province of Bolzano and hosts the permanent exhibition of the mummy Ötzi, “the Iceman”, whereas the Museum of Modern and Contemporary Art (shortened to MART) is placed in the province of Trento and owns one of the most important collections in Italy for what concerns this artistic period.

The article will first proceed by outlining the research objectives, overviewing the clustering technique adopted, presenting the sample and questionnaire employed, and discussing the clustering results combined with binary and multiple Logit analysis. Both academic and practical implications, limitations of the research and future perspectives are provided.

## **2. Research Objective**

The focus of this paper is to find and describe groups of visitors with similar motivational characteristics in visiting an archaeological and a modern and contemporary art museum. This work constitutes a first attempt in studying whether there can be detected heterogeneous profile of visitors in two different types of museums.

The set of motivational factors for segmentation were measured as binary variables, i.e. “Yes/No”, in the *ad-hoc* survey used in this study. When binary data are used for the sake of clustering observations, it is a common practice in literature to use one of the following approaches: applying a hierarchical clustering method using a dissimilarity measures, such as Jaccard, Russell/Rao, Matching, or Dice, computed on the original data (Řezanková, 2009; Finch, 2005); applying the *k*-means method using the Euclidean measure on the original data (Leisch, 2006); transforming the binary variables into continuous ones through a Factor Analysis, like the Correspondence Analysis, and then use the results as input of a clustering method using the Euclidean distance (Bouguila, 2010). When *k*-means is applied on the original data, the centres give the conditional marginal probabilities of observing a “1” (i.e., “YES” in one of the segmentation variables) given the cluster membership. It is important to underline that the dissimilarity measures for binary data have been extensively used and analysed with hierarchical methods but not with partitioning one, in which these types of measures are less common. In this context the Bagged Clustering method can be viewed as a useful solution for two reasons: it allows segmenting visitors by using the original binary data, and it overcomes the main limitations of the traditional segmentation methods.

### 3. Methodology

In this study, the Bagged Clustering method proposed by Leisch (1999) was adopted. This method is a combination of partitioning and hierarchical procedures and consists of the following steps:

1. First of all,  $B$  bootstrap sample  $X_N^1, \dots, X_N^B$  were constructed by drawing with replacement from the original sample  $X_N$ , where  $N$  is the sample size.
2. A partitioning method, called base method, is chosen by the researcher (e.g.,  $k$ -means) and applied to each bootstrapped sample. From this procedure,  $B \times K$  centres  $c_1^1, \dots, c_K^1, c_1^2, \dots, c_K^2, c_1^B, \dots, c_K^B$  are obtained, where  $K$  is the number of centres used in the base clustering method and  $c_j^i$  is the  $j$ th centre ( $j = 1, \dots, K$ ) of  $X_N^i$ , which is the  $i$ th bootstrap sample ( $i = 1, \dots, B$ ).
3. All the centres are combined into a new dataset  $C_{B \times K}$ .
4. A hierarchical cluster algorithm is applied to the  $C_{B \times K}$  dataset in order to produce a partition of the centres.
5. The final outcome is displayed through the usual dendrogram of classical hierarchical methods, where the best partition of centres is obtained by simply investigating it. Finally, the partition of the original observations results from assigning the  $x \in X_N$  observations to their closest centres. In this way each observation is assigned to the cluster containing the centre to which it is associated.

Figure 1 schematically represents the steps that characterize the Bagged Clustering.

<INSERT FIGURE 1 BY HERE>

This method can be interpreted as both a complexity-reducing pre-processing stage for the hierarchical methods and a combination procedure of several partitioning results (Kang, Zhang, & Fan, 2008; Leisch, 1999). It has a better performance in comparison to other standard clustering methods for both continuous and binary data sets (Leisch, 1999). Furthermore, the Bagged Clustering technique overcomes many limitations of both partitioning and hierarchical algorithms. Partitioning methods are more flexible and perform better with large dataset than hierarchical methods (Everitt et al., 2011). The latter have the disadvantage that once observations are merged with others in a group, they cannot be removed from that cluster. However, many partitioning algorithms depend strongly on the

starting selected centres because they are based on iterative stochastic procedures. Thus running the  $k$ -means algorithm twice on the same dataset with different starting centres may result in two different solutions, for the less clear the hidden data structure, the higher the difference between the two solutions. From this point of view  $k$ -means is an unstable algorithm, though widely used. The reason is related to the possibility of finding at each run only a local and unstable solution rather than a global one (Jain, 1999). In addition, the latter can be absent. The Bagged Clustering technique instead is more stable and it is less dependent on the starting solution. In fact, running the base method (i.e. the partitioning methods chosen) on  $B$  bootstrapped sample is equivalent to running  $B$  times the base method, and then obtaining the final solution from summarizing all these results properly.

Another limitation in using  $k$ -means is that it is necessary to select the number of clusters in advance (Buttrey & Karo, 2002; Jain, 1999). In tourism studies, while using non-hierarchical algorithm it is common practice to decide the number of clusters on the basis of practical and subjective preferences (Choi, 2011; Konu, Laukkanen, & Komppula, 2011; Albalade & Bel, 2010; Pérez & Nadal, 2005) or derive this information from applying a hierarchical cluster method (Claver-Cortés, Molina-Azorín, & Pereira-Moliner, 2007; Bigné & Andreu, 2004; Chen & Hsu, 1999; Punj & Steward, 1983). Although many internal validity indices were developed in order to overcome this problem and to drive the researchers to select this number properly (see for example Handl, Knowles, & Kell, 2005), none has yet been globally accepted, and in the tourism field they have not been widely applied (see Brida, Disegna, & Osti 2012 for an example of application). Furthermore, as underlined by Vesanto & Alhoniemi (2000), in practice the value of these indices must be interpreted as a mere guideline. The use of the Bagged Clustering algorithm overcomes also the problem of selecting the “optimal” number of groups. Although an initial choice of a number of groups is required, it does not affect the final results. The “final” number of clusters is in fact obtained *a posteriori* as a result of the hierarchical algorithm.

Finally, only when binary data is used the interpretation of the results from Bagged Clustering is easier and more exhaustive than traditional methods. In fact, a  $k$ -means algorithm applied to a binary dataset produces  $K$  centres, one per each segment. Each centre is a  $d$ -dimensional vector (where  $d$  denotes the number of variables used for segmentation) in which each value can be interpreted only in term of the frequency (i.e. mean value) with which the value 1 occurred among the units observed in the selected segment for each variable. Each segment created by the Bagged Clustering algorithm is composed by several

$d$ -dimensional centres. Therefore the empirical distribution of the centres can be easily checked and represented through the standard Box-plot for each variable and segment.

## **4. Data and structure of the questionnaire**

### *4.1 The museums*

The research involved the two main museums of Trentino-South Tyrol, an Italian region located in the North-East. The ÖTZI museum is situated in Bolzano, the main city of South Tyrol. It hosts the permanent exhibition a mummy from the Neolithic period Ötzi. The mummy was discovered in the region in 1991 as one of the oldest mummies in the world. The good preservation status of it and its belongings has attracted researchers and visitors from around the world, and made the museum the most important cultural attraction in Bolzano.

The MART museum, that is the second museum under consideration, is placed in Trento and Rovereto, the two main cities of the province of Trento. It hosts both a permanent collection of modern art, where works are displayed on a rotating basis, and a temporary exhibition. It has the most important collections in Italy of modern and contemporary art works, in particular futurism. As pointed by Brida, Pulina, and Meleddu (2012), the idea of a museum for modern and contemporary art was born in the late 1970s, against the background of industrial and unemployment crisis.

### *4.2 Research design*

The research is based on a survey conducted from June to September 2011 among the visitors of the ÖTZI and MART museums. A total of 1,288 interviews were successfully collected almost equally divided between MART (46%) and ÖTZI (54%). 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 for testing the questionnaire before conducting the full survey, in order to avoid biases related to its structure and wording.

Interviews were held to visitors exiting the museums after their visit, in selected working and weekend days of the four months analysed, and during different times of the day. Only one person per travel party was selected. The questionnaires were anonymous and self-administered in three languages (Italian, German and English), though a research team

member was present to respond if questions or doubts emerged. A convenience sampling method (Cochran, 1977) was used, as there was no sufficient information on the characteristics of museums visitors in order to apply a probabilistic design. The sections of the questionnaire are presented schematically in Table 1.

**<INSERT TABLE 1 BY HERE>**

## **5. Discussion of the results**

As reported above, the set of variables for segmenting interviewees referred to their motivations to visit each of the two museums. The questionnaire asked the respondents if they agreed or not (dichotomous answer) with a set of push factors that motivated the visit to the museum. The set of factors included: satisfying a curiosity (“curiosity”), resting/relaxing (“relax”), a specific interest in such an attraction (“interest”), accompanying friend/family member with a specific interest in such an attraction (“friend”), learning something new (“learn”), telling friends about the visit (“tell”), doing something that one ought to do (“do”), contributing to preserving this attraction for future generation (“future”), revisiting this museum (“revisit”), showing the museum to friends or relatives (“show”), professional or academic reasons (“work”), doing something worthwhile (“worthwhile”), occupying some leisure time (“leisure”), visiting temporary exhibition (“temporary”), seeing the building (“building”).

The Bagged Clustering algorithm considered the  $k$ -means as base method, with  $K=20$  centres and 10,000 iterations was used as base method. A number of  $B=50$  bootstrap samples were considered, resulting in a total of 1,000 ( $B \times K$ ) centres, which were then hierarchically clustered using Euclidean distance and Ward’s agglomerative linkage method. These parameters were chosen because they provided the best performances in previous studies, which used simulated artificial datasets with similar characteristics to the one of this paper (Dolnicar & Leisch, 2004, 2000). Results of the Bagged Clustering method for the MART analysis are graphically showed in Figures 2 and 4, whereas Figures 3 and 5 display those of ÖTZI.

**<INSERT FIGURE 2 BY HERE>**

**<INSERT FIGURE 3 BY HERE>**

The top part of the graph in Figures 2 and 3 display the dendrograms deriving from the procedure, respectively for MART and ÖTZI. The plot under each dendrogram shows the distance of aggregation for each cluster, where the black line reports standardized absolute heights and the grey one stands for first differences. The accentuated bend in the grey line suggests that for MART the correct number of clusters is three, whereas for ÖTZI it is two. These correspond to cutting the dendrogram where the longest distance between two consecutive aggregations appear.

**<INSERT FIGURE 4 BY HERE>**

**<INSERT FIGURE 5 BY HERE>**

The Box-plots in Figures 4 and 5 report the distribution of the centers for each segmentation variable and within each group. In addition the red line that runs across all the Box-plot indicates the sample mean, i.e. the average frequency of “Yes” for each motivational factor. For the sake of interpretation, it is important to emphasize that the higher the height of the box (i.e. the 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 a strong dispersion within a variable indicates non homogeneity of the units of the segment with respect to that characteristic.

The Chi-square test on the centres was calculated in order to test the null hypothesis of independence of each motivation variable from the observed segment. MART results revealed that only telling friends about the visit at the museum (“tell”) is independent from the segments identified ( $p$ -value = 0.744), whereas all other ones significantly depend from the observed segments ( $p$ -value < 0.01). On the other hand ÖTZI reported that only the following motivational factors depended significantly from the identified segments ( $p$ -value < 0.01): satisfying a curiosity (“curiosity”), resting/relaxing (“relax”), accompanying friend/family member with a specific interest in such an attraction (“friend”), learning something new (“learn”), telling friends about the visit (“tell”), doing something that one ought to do (“do”), doing something worthwhile (“worthwhile”), and occupying some leisure time (“leisure”). The Box-plots of MART in Figure 4 revealed the presence of a niche segment (cluster 3, which contains 7.6% of the whole MART sample) and two segments of

almost equal size (42.8% and 49.6% of the total visitors of MART respectively in the cluster 1 and 2). Visitors of cluster 3 visited the museum mainly because for the sake of satisfying a curiosity (“curiosity”), learning something new (“learn”) and doing something worthwhile (“worthwhile”). Therefore, these visitors have been named “Knowledge seeker”. Visitors of cluster 2, named “Interested”, seemed to be strongly attracted by the temporary exhibition (“temporary”) and by a specific interest in such an attraction (“interest”). Cluster 1 collected instead all the remainder visitors. They had in common the fact that they declared that their motivation in visiting MART was “not” one of those considered. Moreover, compared to the other two groups, cluster 1 exhibits lower median values in the majority of the factors, which implies that a big part of these visitors did not select the motivation items proposed in the questionnaire. In particular, it resulted that they did “not” visit MART because it is something that one ought to do (“do”), or to contribute to preserve this attraction for future generation (“future”), revisiting this museum (“revisit”), doing something worthwhile (“worthwhile”), occupying some leisure time (“leisure”). Therefore, cluster 1 is labelled “Non-motivated”.

Clusters of ÖTZI visitors recall similar segments as those identified for MART. About 25% of respondents are grouped in cluster 1 where the main motivations are satisfying a curiosity (“curiosity”) and learning something new (“learn”), such as the “Knowledge seeker” visitors identified as niche in the MART dataset. If “Knowledge seekers” represented a niche for MART, this segment is bigger for ÖTZI. The remainder 75% was clustered into the “Non-motivated” cluster similarly to MART dataset. This cluster groups visitors that did not report any of the proposed motivations. In particular, they did not visit ÖTZI for the sakes of resting/relaxing (“relax”), telling friends about the experience (“tell”), or, as for MART’s similar group, doing something that one ought to do (“do”), doing something worthwhile (“worthwhile”), occupying some leisure time (“leisure”).

### *5.1. Clusters description*

The additional information collected through the survey were used to characterize the clusters identified in each museum in terms of socio-demographic (gender, age, level of education, origin, occupation, and visiting party) and economic (household income, total expenditure per person per night, and expenditure at the shop of the museum per person) variables. Table A1 in the Appendix A reports the complete list of these profiling variables with a brief description of them.

**<INSERT TABLE 2 BY HERE>**

Some statistically significant dependency emerged between clusters and the profiling variables for each museum (Table 2). Among the visitors of MART, the “Interested” was on average older (47 years old) and mainly employed, whereas the “Knowledge seeker” was of younger age (39 years old on average) though it reports the highest percentage of retired. “Non-motivated” was instead more in an autonomous worker or in other position than the remainder.

The origin of the visitors was main distinctive variable between the two clusters for the ÖTZI museum. The “Knowledge seeker” came mainly from abroad (Germany or other cities), whereas “Non-motivated” were mainly Italian. This result was not surprising because the marketing policy of this museum is actually more oriented to the foreign countries than to Italy. Furthermore, the National Geographic Magazine dedicated it articles written in English or German, thus contributing to promote this museum in a good way.

**<INSERT TABLE 3 BY HERE>**

### *5.2. Profiling tourists: Logit models*

Logit models were then estimated in order to assess whether the set of socio-demographic and economic characteristics was a significant predictor of the likelihood to be part of one of the groups. Due to the presence of 3 groups for MART and 2 for ÖTZI, Multinomial Logit and Logit models were respectively adopted. For both museums the “Non-motivated” cluster was used as baseline. Table 3 reports the coefficients for the resulting models. MART’s Multinomial Logit result leads to the conclusion that “Interested” visitors were significantly discriminated from the “Non-motivated” group only because they were willing to spend more at the shop of the MART museum. It is relevant to remind that “Interested” visitors were highly interested in the temporary exhibition (“temporary”) and in the museum as attraction (“interest”). Therefore, from these empirical results one can derive that they were likely to be interested in spending a higher amount of money at the shop, and probably buying souvenirs like books that were the most expensive ones. “Knowledge seeker” were instead more likely to be male, come from the Northern Italy, and with higher levels of education (university degree or postgraduate), whereas in comparison to them the “Non-motivated” cluster was likely to come from abroad. A higher education of knowledge seekers is in line with the main

motivations that push these visitors to make the visit: curiosity, learning and doing something worthwhile.

The age, the origin, and the visiting party were instead the variables that significantly discriminate the “Knowledge seeker” from the “Non-motivated” of ÖTZI. The lower the age of the visitors, the higher (less than proportionally) the probability of being a “Knowledge seeker”. Foreign visitors were more likely to be grouped in the “Knowledge seeker” and the members of this group of visitors were more likely to visit ÖTZI alone or with their families, instead in groups of people that were not member of their families.

## 6. Conclusions

The Bagged Clustering method proposed by Leisch (1999) was adopted in this research as a useful exploratory technique and a suitable alternative to the more traditional clustering methods. In particular, this method doesn't require the *a priori* definition of the number of clusters and the result is more stable than those obtain by a more traditional partitioning method, like the *k*-means. Even if dichotomous variables are used, this method allows using the Box-plots to easily represent the empirical distribution of each segmenting variable among the clusters identified. Furthermore, in contrast with other segmentation procedures commonly used in the tourism literature for binary datasets (like the use of the factor analysis followed by a hierarchical cluster method), the Bagged Clustering allows considering the original set of variables using all available information. As underlined by Chen and Hsu (1999), there are two main reasons to use all variables for segmenting purposes: firstly, results of the factor analysis could be different when researcher uses different rotation methods; secondly, the original variables may be more interpretable than derived constructs with factor labels because the naming and the interpretation of the construct involve personal judgments that it is generally preferred to avoid.

This algorithm was presented in this research both from a theoretical and an applied point of view. The dataset used were collected from two ad-hoc surveys conducted from June to September 2011 at the two main museums of the two provinces of the Trentino-South Tyrol region (Trento and Bolzano): the South Tyrol Museum of Archaeology (shortened to “ÖTZI”) and the Museum of Modern and Contemporary Art (shortened to “MART”).

The aim was to segment the visitors of these two different types of museums with respect to the motivational factors that push them to make the visit. Furthermore, a comparison between the cluster solutions obtained for the two museums was developed.

The visitors of MART museum were grouped into three clusters while those of the ÖTZI museum into two clusters. In both cases:

- A group of “Knowledge seeker” emerged, having heterogeneous socio-demographic and economic characteristics between the museums. In fact, the “Knowledge seeker” of MART was a group of visitors living near the museum (in Northern Italy) with high level of education, whereas the “Knowledge seeker” of ÖTZI was a group of young, foreign visitors who preferred to visit the museum with their families or alone. Therefore, MART museum must driving its promotional and marketing efforts in order to attract highly educated and foreign visitors. Vice versa, the promotional policy of ÖTZI museum should be more oriented to capture the Italian visitors, mainly young people or families.
- A large group of visitors with any particular push-motive (the “Non-motivated”) was found in each museum and must be analysed more in depth in future researches.

The last, and largest, group identified among the MART visitors, i.e. the “Interested”, seemed to spend more than the other groups at the shop of the museum. In fact, they visited the museum because they are interested in such an attraction and in visiting the temporary exhibition. These two push factors can lead to buy more books, or other expensive souvenirs, than other visitors. So that, the museum should propose more books or souvenirs regarding its exhibitions and its attractions in order to increase the revenues of the museum shop.

The main limitation of this study is that the segmentation analysis is based on a non-random sampling technique. Thus, to verify if the results of this research are valid for other museums, a future study will be required in other museums, in other years, and/ or other towns. Future researches must focus on the looking for a more suitable distance measure for binary data to apply both in the base method and in the hierarchical analysis, in order to improve the Bagged Clustering method.

Finally, this research could be extended in the future by comparing the Bagged Clustering results with other segmentation techniques, both linear and non-linear techniques, evaluating which of them is better for segmenting visitors of museums and cultural attractions.

## References

- Albalade, D., & Bel, G. (2010). Tourism and urban public transport: Holding demand pressure under supply constraints. *Tourism Management*, 31(3), 425–433.
- Beane, T. T., & Ennis, D. M. (1987). Market segmentation: A review. *European Journal of Marketing*, 21(5), 20–42.
- Bennett, T. (1994). *The reluctant museum visitor: A study of non-goers to history museums and art galleries*. Sydney: Australia Council.
- Bigné, J. E., & Andreu, L. (2004). Amotions in Segmentation. An Empirical Study. *Annals of Tourism Research*, 31(3), 682–696.
- Bouguila, N. (2010). On multivariate binary data clustering and feature weighting. *Computational Statistics and Data Analysis*, 54, 120-134.
- Brida, J. G., Disegna, M., & Osti, L. (2012). 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., Pulina, M. and Meleddu, M. (2012). Factors influencing the intention to revisit a cultural attraction: the case of MART of Rovereto. *Journal of Cultural Heritage*, 13(2), 167–174.
- Budayan, C., Dikmen, I., & Birgonul, M.T. (2009). Comparing the performance of traditional cluster analysis, self-organizing maps and fuzzy C-means method for strategic grouping. *Expert System with Application*, 36, 11772-11781.
- Buttrey, S. E., & Karo, C. (2002). Using K-nearest-neighbor classification in the leaves of a tree. *Computational Statistics and Data Analysis*, 40, 27-37.
- Chen, J. S., & Hsu, C. H. C. (1999). The use of logit analysis to enhance market segmentation methodology. *Journal of Hospitality & Tourism Research*, 23(3), 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(3), 727–745.
- Cochran, W. G. (1977). *Sampling techniques*. Wiley series in probability and mathematical statistics (Third Edn.). John Wiley & Sons.
- Dicks, B. (2003). *Culture on display: The production of contemporary visitability*. Berkshire: Open University Press.

- 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(3), 281–292.
- Dolnicar, S., & Leisch, F. (2004). Segmenting Markets by Bagged Clustering. *Australasian Marketing Journal*, 12(1), 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(1), 44–52.
- Everitt, B.S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis*. Wiley series in Probability and Statistics, London.
- Fingh, H. (2005). Comparison of distance measures in cluster analysis with dichotomous data. *Journal of Data Science*, 3, 85-100.
- Handl, J., Knowles, J., & Kell, D. B. (2005). Computational Cluster Validation in Post-Genomic Data Analysis. *Bioinformatics*, 21(15), 3201–3212.
- Huang, S. C., Chang, E. C., & Wu H. H. (2009). A case study of applying data mining techniques in an outfitter’s customer value analysis. *Expert System with Application*, 36, 5909-5915.
- Hughes, H. L. (2002). Culture and tourism: A framework for further analysis. *Managing Leisure*, 7(3), 164–175.
- Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data Clustering: A review. *ACM Computing Surveys*, 31(3), 264-323.
- 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.
- Kim, H., Cheng, C. K., & O’Leary, T. J. (2007). Understanding participation patterns and trends in tourism cultural attractions. *Tourism Management*, 28(5), 1366–1371.
- Konu, H., Laukkanen, T., & Komppula, R. (2011). Using ski destination choice criteria to segment Finnish ski resort customers. *Tourism Management*, 32(5), 1096–1105.
- Kotler, P., Bowen, J. T., & Makens, J. C. (2010). *Marketing for hospitality and tourism* (5th Edn.). Upper Saddle River, New Jersey: Pearson Prentice Hall.

- Leisch, F. (1999). Bagged Clustering. *Working paper 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.
- Lia, S. H., Chu, P. H., & Hsiao P. Y. (2009). Data mining techniques and applications – A decade review from 2000 to 2011. *Expert System with Application*, 39, 11303-11311.
- MacDonald, G., & Alford, S. (1995). Canadian Museums and the Representation of Culture in a Multicultural Museum. *Cultural Dynamics*, 7(1), 15–36.
- McKercher, B. (2004). A comparative study of international cultural tourists. *Journal of Hospitality and Tourism Management*, 11(2), 95–107.
- Pérez, E. A., & Nadal, J. R. (2005). Host Community Perceptions. A Cluster Analysis. *Annals of Tourism Research*, 32(4), 925–941.
- Punj, G., & Steward, D. W. (1983). Cluster analysis in marketing research: Review and suggestions for application. *Journal of Marketing Research*, 20(2), 138–148.
- Řezanková, H. (2009). Cluster analysis and categorical data. *Statistika*, 3, 216-232.
- Saarenvirta, G. (1998). *Mining customer data*. DB2 Magazine, 3(3), 10–20.
- Schuster, M. D. (1991). *The audience for American art museums. National endowment for the arts research division report 23*. Washington: NEA.
- Stylianou-Lambert, T. (2011). Gazing from home: cultural tourism and art museums. *Annals of Tourism Research*, 38(2), 403–421.
- Vesanto, J., & Alhoniemi, E. (2000). Clustering of the self-organizing map. *IEEE Transactions on Neural Networks*, 11(3), 586–600.

Figure 1. The bagged clustering algorithm.

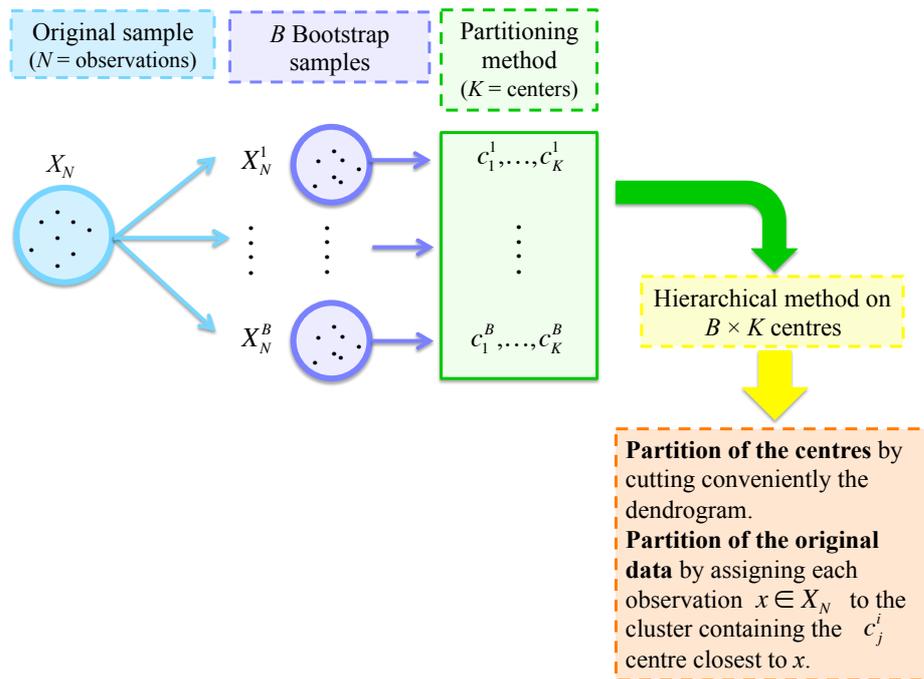


Figure 2. MART museum: Bagged Clustering dendrogram together with the plot regarding the relative height of aggregation (black line) and the first differences (grey line).

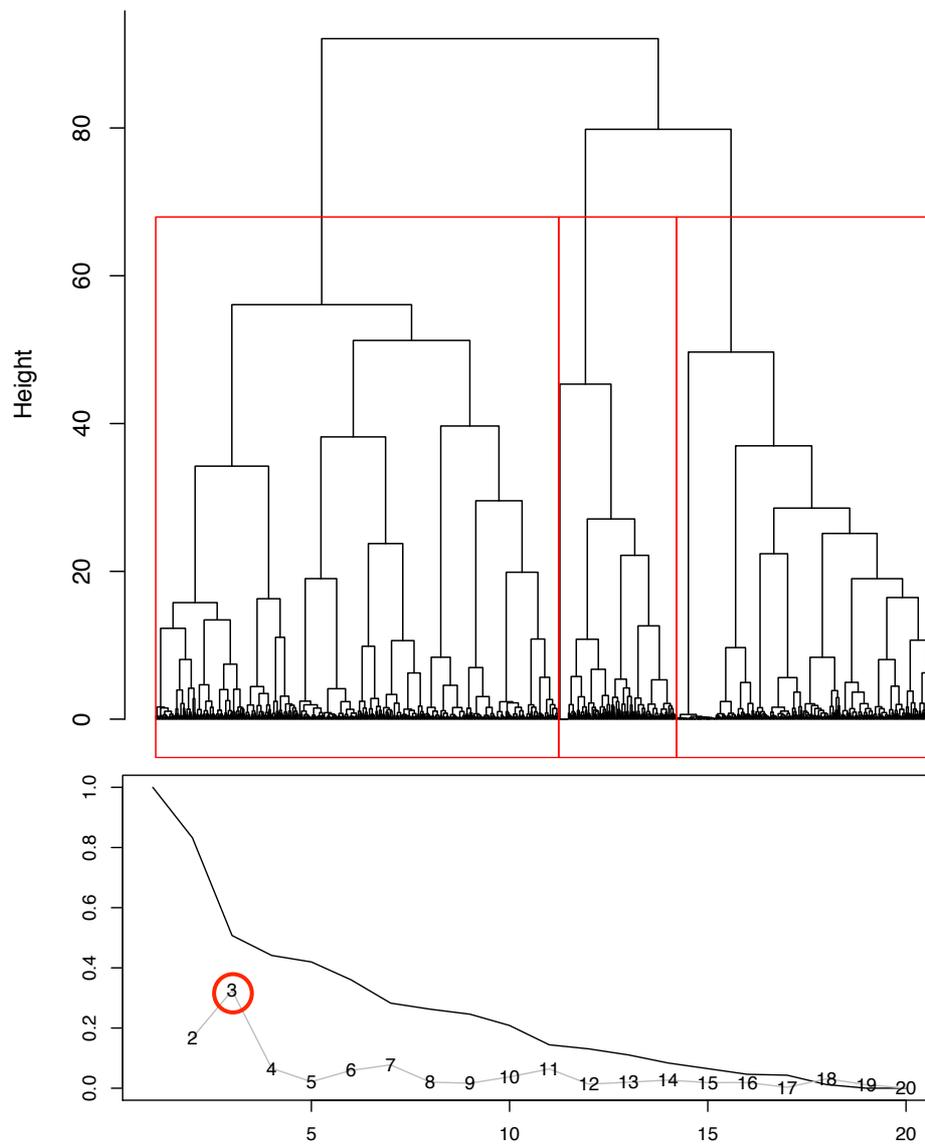


Figure 3. ÖtZI museum: Bagged Clustering dendrogram together with the plot regarding the relative height of aggregation (black line) and the first differences (grey line).

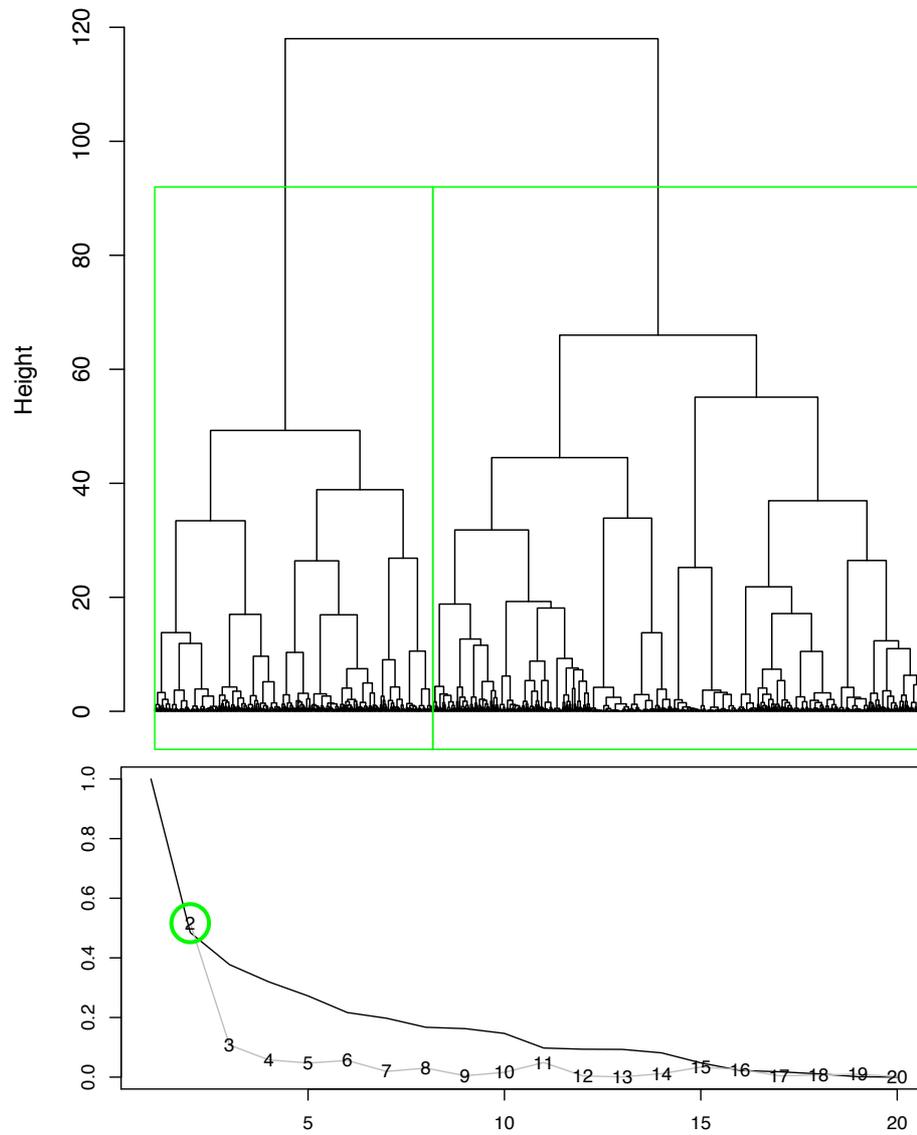
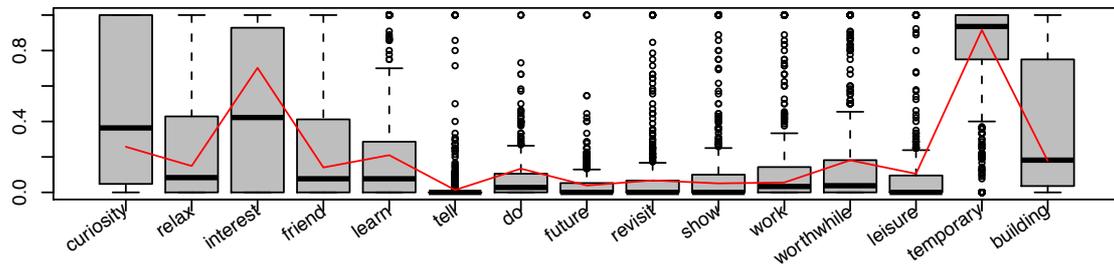
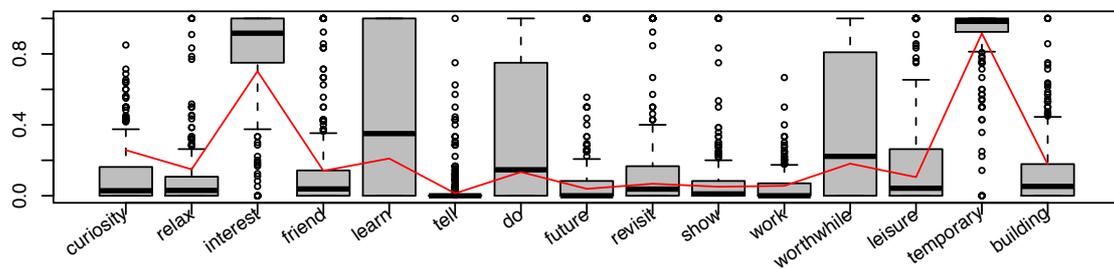


Figure 4. MART museum: box-plot for the three clusters solution.

**Cluster 1: 517 centers, 253 data points**



**Cluster 2: 332 centers, 293 data points**



**Cluster 3: 151 centers, 45 data points**

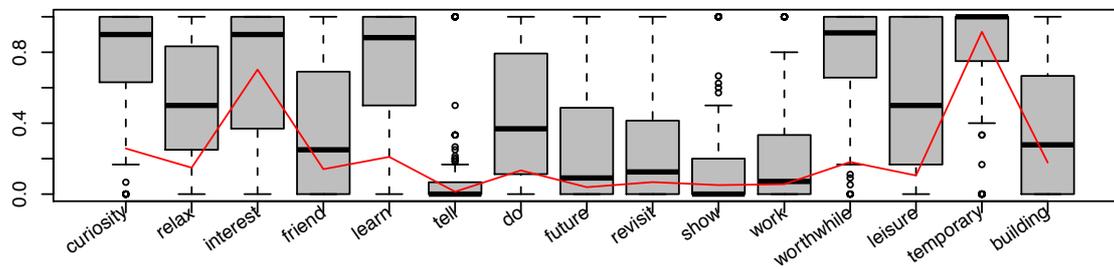
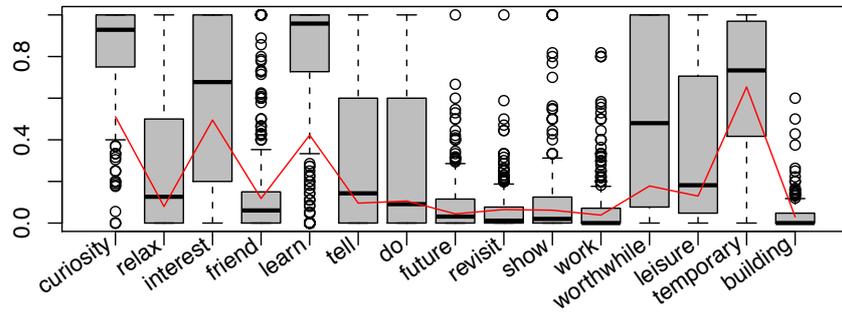


Figure 5. ÖTZI museum: box-plot for the two clusters solution.

**Cluster 1: 361 centers, 173 data points**



**Cluster 2: 639 centers, 506 data points**

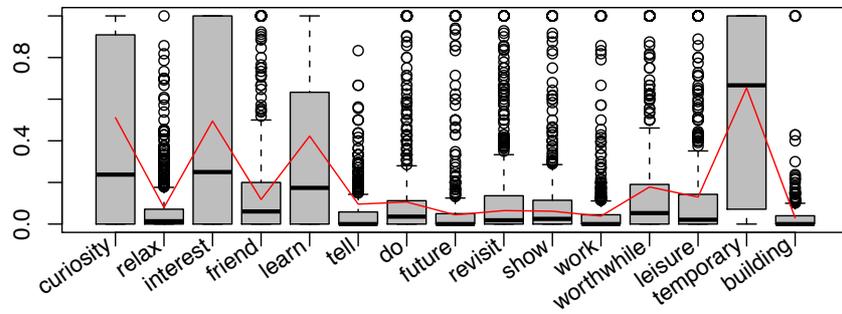


Table 1. Structure of the questionnaire.

Sections	Description	Categories of variables
I	Museum information	Repeat visiting; number of museums visited in the last year; factors that stimulated the visit <sup>*</sup> ; rating of factors that describe the visit <sup>**</sup> ; shopping expenditure at the museum; authenticity perception <sup>*</sup> .
II	Trip information	Motives of the trip; number of nights, total expenditure per person per night.
III	Interviewees' profile	Some socio-demographic and economic characteristics of interviewees and their families.

Notes: <sup>\*</sup> dichotomous variables have been used; <sup>\*\*</sup> A 5-points Likert scale has been used.

Table 2. Socio-demographic and economic characteristics.

	MART				ÖTZI		
	Interested	Knowledge seeker	Non-motivated	<i>p</i> -value	Knowledge seeker	Non-motivated	<i>p</i> -value
<i>Male (%)</i>	41.30	55.56	45.63	0.17	53.66	50.61	0.49
<i>Age (mean)</i>	46.94	39.38	42.07	***	44.58	44.12	0.68
<i>University (%)</i>	81.57	93.33	82.14	0.14	68.48	68.74	0.95
<i>Origin of visitors (%)</i>				0.41			***
Abroad	3.08	0.00	3.97		25.60	16.09	
Germany	3.42	2.22	4.37		42.86	31.57	
Centre/South of Italy	7.19	11.11	11.51		5.95	17.92	
North-East of Italy	48.97	60.00	41.67		8.33	15.68	
North-West of Italy	13.70	11.11	14.68		14.88	14.46	
Local resident	23.63	15.56	23.81		2.38	4.28	
<i>Occupation (%)</i>				**			0.94
Autonomous worker	17.41	13.33	18.65		21.95	19.96	
Employed	49.83	37.78	46.83		58.54	60.29	
Retired	15.70	37.78	9.52		12.20	12.89	
Other occupations	17.06	11.11	25.00		7.32	6.86	
<i>Visiting party (%)</i>				0.85			0.31
Alone	8.19	6.67	7.91		8.33	5.69	
Couple	12.97	11.11	34.39		41.67	36.79	
Children	35.84	28.89	15.42		33.93	40.24	
Group	43.00	53.33	42.29		16.07	17.28	
<i>Household annual income (%)</i>				0.34			0.14
0 -  25,000	19.45	15.56	20.63		5.45	10.86	
25,000 -  50,000	40.27	51.11	36.11		27.88	26.64	
50,000 -  75,000	9.90	15.56	13.49		13.33	16.19	
> 75,000	7.51	0.00	8.33		20.00	14.55	
Missing income	22.87	17.78	21.43		33.33	31.76	
<i>Expenditure (mean)</i>							
Total expenditure	30.38	23.62	33.42	0.54	55.21	48.66	0.13
Shopping at the museum	13.78	8.18	9.18	0.12	5.55	10.08	0.48

Notes:

Chi-square test was used for qualitative variables and continuous variables recoded in classes. ANOVA test and t-test were used in order to test whether the mean value of the quantitative variables significantly differ among the three clusters identified for the MART museum and between the two clusters identified for the ÖTZI museum.

All test results are not significant unless indicated otherwise: \*\*\*Significant at  $p \leq 0.01$ , \*\*Significant at  $p \leq 0.05$ ,

\*Significant at  $p \leq 0.1$

Table 3. Multinomial Logit (MART) and Binary Logit (ÖTZI) coefficients.

	MART <sup>A</sup>		ÖTZI <sup>B</sup>
	Interested	Knowledge seeker	Knowledge seeker
<i>Male</i>	-0.25 (0.19)	0.66 (0.33)**	0.05 (0.20)
<i>Age</i>	>0.01 (0.06)	-0.01 (0.08)	-0.13 (0.06)**
<i>Age2</i>	>0.01 (>0.01)	-0.01 (>0.01)	>0.01 (>0.01)**
<i>University</i>	-0.01 (0.24)	1.41 (0.61)**	0.14 (0.22)
<i>Origin of visitors</i>			
Abroad	-0.50 (0.50)	-29.79 (0.66)***	1.05 (0.59)*
Germany	-0.09 (0.48)	0.63 (1.18)	0.94 (0.56)*
Centre/South of Italy	-0.50 (0.38)	0.86 (0.70)	-0.51 (0.65)
North-East of Italy	0.11 (0.23)	1.06 (0.46)**	0.03 (0.62)
North-West of Italy	-0.10 (0.32)	0.63 (0.63)	0.71 (0.60)
<i>Occupation</i>			
Autonomous worker	0.16 (0.32)	-0.66 (0.61)	0.34 (0.35)
Employed	0.36 (0.28)	-0.65 (0.44)	0.21 (0.30)
<i>Visiting party</i>			
Alone	-0.03 (0.35)	-0.51 (0.71)	0.74 (0.43)*
Couple	-0.12 (0.21)	-0.25 (0.39)	0.52 (0.26)**
Children	-0.32 (0.29)	-0.28 (0.57)	0.50 (0.28)*
<i>Household annual income</i>			
Income	-0.01 (>0.01)	-0.01 (0.01)	>0.01 (>0.01)
Missing income	-0.07 (0.30)	-0.88 (0.51)*	0.26 (0.30)
<i>Expenditure</i>			
Total expenditure	-0.01 (>0.01)	-0.01 (0.01)	>0.01 (>0.01)
Missing total expenditure	-0.01 (0.21)	-0.21 (0.38)	-0.23 (0.25)
Shopping at the museum	0.03 (0.02)**	0.01 (0.03)	-0.03 (0.03)
Missing expenditure at the shop of the museum	-0.32 (0.21)	0.32 (0.37)	-0.15 (0.24)
<i>Constant</i>	-0.33 (1.12)	-2.44 (1.75)	0.21 (1.22)

*Notes:*

All test results are not significant unless indicated otherwise: \*\*\*Significant at  $p \leq 0.01$ , \*\*Significant at  $p \leq 0.05$ ,

\*Significant at  $p \leq 0.1$ . Robust Std. Err. in brackets.

<sup>A</sup> Multinomial logit:  $N = 588$ ; Wald  $\chi^2(40) = 9207.03$ ; Prob >  $\chi^2 = 0.00$ ; Pseudo  $R^2 = 0.0655$ ; McFadden  $R^2 = 0.066$ ; Cox & Snell  $R^2 = 0.112$ ; Nagelkerke  $R^2 = 0.134$ .

<sup>B</sup> Binary Logit:  $N = 631$ ; Wald  $\chi^2(20) = 40.40$ ; Prob >  $\chi^2 = 0.00$ ; Pseudo  $R^2 = 0.0626$ ; McFadden  $R^2 = 0.063$ ; Cox & Snell  $R^2 = 0.068$ ; Nagelkerke  $R^2 = 0.101$ .

## Appendix A

Table A1. Description of the explanatory variables.

<b>Independent variables</b>	<b>Descriptions</b>
<i>Socio-demographic and economic characteristics</i>	
Male	1= male; 0= female
Age	Age of the respondent (continuous)
Age2	Squared age of the respondent (continuous)
University	1 = education level is university degree or postgraduate; 0= otherwise
<i>Origin of visitors</i>	
Abroad	1= Abroad (excluding Germany); 0= otherwise
Germany	1= Germany; 0= otherwise
Centre/South of Italy	1= Centre, South, or islands of Italy; 0= otherwise
North–East of Italy	1= North-East of Italy (excluding the province in which the museum is located); 0= otherwise
North–West of Italy	1= North-West of Italy; 0= otherwise
Local resident	1= province that host the museum; 0= otherwise (reference category)
<i>Occupation</i>	
Autonomous worker	1= autonomous worker; 0= otherwise
Employed	1= employed (full-time or part-time); 0= otherwise
Retired	1= retired; 0= otherwise
Other occupation	1= student/unemployed/housewife/working occasional or on project/teacher/other; 0= otherwise (reference category)
<i>Visiting party</i>	
Alone	1= alone; 0 = otherwise
Couple	1= partner/spouse; 0 = otherwise
Children	1= children between 0 and 12 years; 0 = otherwise
Group	1= friends/colleagues/organized group/other relatives; 0 = otherwise (reference category)
<i>Household annual income</i>	
Income	Central value of each income category (see the list reported in Table 2); 0 if the respondent does not declare her income (continuous)
Missing income	1 = respondent does not declare his/her income; 0 = otherwise
<i>Expenditure</i>	
Total expenditure	Individual expenditure for accommodation, food and beverage, shopping in the shops of the city, pharmacy, tour guide services, other expenditures linked to the visit (excluding expenditure for transportation) per night in Euros; 0 if respondent does not state their expenditure (continuous)
Missing total expenditure	1 = respondent does not declare his/her total expenditure; 0 = otherwise
Shopping at the museum	Individual expenditure at the shop of the museum in Euros; 0 if respondent does not state their expenditure (continuous)
Missing expenditure at the shop of the museum	1 = respondent does not declare his/her expenditure at the shop of the museum; 0 = otherwise