

Application of Machine Learning to Slow Tourism Market Segmentation: A Case Study at Nanzhuang

Ming-Gong Lee* and Che-Chia Nien

Department of Tourism and Leisure/Ph.D. Program in Engineering Science, Chung Hua University

* mglee1990@gmail.com

Graduate Student, Department of Computer Science and Information Engineering, Chung Hua University

Abstract: Truly little is known about the characteristics of tourists to Nanzhuang township, one of four Cittaslow townships in Taiwan. This study gives implication about the distinctive styles of tourists to Nanzhuang. Today, Nanzhuang township is famous for its role as a slow city in Taiwan. It is reasonable to assume this characteristic would have influence on its tourism market. In order to understand types of tourism to Nanzhuang and to apply a beneficial advertising strategy in terms of the principles of the Cittaslow organization, we need to know the features of tourists to this township. A survey drafted according to the Cittaslow principles was performed, and a total of 222 responses were collected. Machine learning (ML) tools for data science, such as, *k*-means clustering, Principal Component Analysis (PCA), and one-way ANOVA were applied to do proper clustering and analysis of the data. The results have shown that the tourists can be suitably categorized into three distinct groups: Advocates of Slow Tourism (AST); Conscious of Slow Tourism (CST), and Unconscious of Slow Tourism (UST). Interestingly, the descriptive statistics between these three groups do not show any difference in their background, for instance, regarding age and education. A precise marketing strategy for slow tourism should be carefully considered accordingly.

1. Introduction

Tourism is a so-called Chimneyless Industry; it can bring up the GDP for an entire nation. Prior to the Covid-19 pandemic, the tourism industry accounted for around 4~5% of the Gross Domestic Product (GDP) of Taiwan and has the possibility to increase to around 6~7 % with a surge in the number of incoming tourists. During the Covid-19 pandemic, however, this percentage has reduced, the effects of which have been felt not only in Taiwan, but the world. As we consider how to revive the tourism industry, a crucial element to observe is the current trends that are becoming apparent in this industry. Slow Tourism is one such trend. In Taiwan, the concept of slow tourism is still relatively new, however, in some cases, people are participating in this trend without awareness of their participation. According to [5], Slow Tourism can be defined as the following:

Slow Tourism must follow two essential principles: taking time and attachment to a particular place. Taking time means modification of the daily time relationship, specifically a different perception of nature and living in harmony with a place, its inhabitants, and their culture. The environment is not merely perceived by sight, but by using all five senses. Tourists must be able to change pace, to look rather than to see, to experience the area rather than to endure it.

By such a definition, we see that the ideals of Slow Tourism can be practiced both by those who are conscious of the concept of Slow Tourism, and those who may not have awareness

of this concept yet perform tourism in a way that aligns with these ideals. For example, camping is a popular activity during vacation in Taiwan, and Nanzhuang township has many ideal camping destinations. It is possible that tourists to Nanzhuang have taken advantage of these locations, regardless of their dedication to the practice of Slow Tourism.

In this study, we investigate the city of Nanzhuang, Taiwan, looking at the style and motivation for tourism of the visitors to this township. Specifically, we seek to understand what role Slow Tourism plays in these tourists' motivation for visiting, whether they are fully aware of the concepts of Slow Tourism, have some knowledge of Slow Tourism, but are unaware that they are in fact a slow tourist, or whether they are just spending their leisure time in this location, oblivious to the fact that they are performing Slow Tourism. For the local city government to properly target their advertisement regarding tourism, it is essential that they know the majority type of tourist seeking out this destination. Effective advertising will then greatly benefit the local businesses in Nanzhuang. If most tourists to this city are in fact aware of Slow Tourism, then it would be important to emphasize this concept and the specific activities and policies that reinforce Slow Tourism in Nanzhuang. If not, exposure to the concepts of Slow Tourism and the ways in which the township applies these concepts could further increase the desire of tourists to visit. Regardless of the majority type of tourist visiting Nanzhuang, we intend with this study to shed light on the type of tourism occurring at this destination, allowing for better targeted tourism advertisement for this area. This in turn, will hopefully benefit the local businesses, residents, and government by increasing the rate of tourism to the city.

In the past, such an investigation would be carried out by traditional statistical methods, however, in recent years machine learning has become popular across many domains. We would like to explore the benefits of a machine learning approach carried out in the field of tourism. Our research applies machine learning to a dataset containing responses from tourists to Nanzhuang. Specifically, we utilize two clustering algorithms to decipher whether unsupervised machine learning techniques can aid in our understanding of the types of tourists visiting this township. Furthermore, we investigate the data to understand and recognize any trends linking the visitors to Nanzhuang. The goal of this study is to understand the types of tourists that travel to this township and identify the level of importance of Slow Tourism amongst these tourists. This paper is organized as follows, we will first give a background on Slow Tourism and some related movements. We will then introduce Nanzhuang township and the dataset that was used for our experiments. This will be followed by our experimental methodology and analysis of our experimental results. Finally, we will discuss ideas for further research and give the conclusion of this study.

1.1 Slow Cities in Taiwan

Currently there are four slow cities in Taiwan that are members of the Cittaslow network. They are Dalin township in Chiayi County, Fonglin township in Hualien County, and two townships in Miaoli County, Sanyi and Nanzhuang. While each of these townships is unique in their culture and character, they are all representative of the ideas of the Cittaslow organization and the concept of a slow city. Each of these townships are agricultural and rely heavily on local produce, which contributes to their unique local cuisine. This research focuses on Nanzhuang township, located in the mid-north region of Taiwan. Population in Nanzhuang consists in large part of the Hakka people, and the Saisiyat people and the Atayal people, both of which are Indigenous peoples of Taiwan [2]. This township is a well-known tourist destination in Taiwan,

offering a drastically distinct experience to that of the more urban destinations, such as the capital city, Taipei. Some of the most popular tourist sites include Nanzhuang Old Street, Shenxian Valley (also known as Valley of the Immortals), and Xiangtian Lake [3]. As mentioned previously, camping is also a popular activity for visitors to Nanzhuang. In fact, among the top thirty camping destinations in Miaoli County, seven of these destinations are in Nanzhuang. Many tourists come to visit these popular destinations to shop, enjoy the old-style atmosphere of the township, eat local food, visit traditional temples and other activities, which is certainly a common style of tourism in Taiwan and other countries. In this way, these visitors are able to immerse themselves in the town's local culture, an important aspect of slow tourism, but does this truly constitute conscious Slow Tourism? Are these visitors aware of their participation in this style of tourism? We would like to understand the awareness amongst these tourists regarding Slow Tourism. To what extent are they seeking out such a destination because of its application of Slow Tourism?

2. Methodology:

2.1 Tourist Dataset

This research was obtained through a questionnaire that consists of two parts: the first part contains fifteen statements that are designed according to the Cittaslow Charter [1] and are used to clarify if there are distinct characteristics of tourists to Nanzhuang, based on their opinion regarding the concepts of the Cittaslow Charter. The questions are adapted from a previous study on Slow Tourism to the city, Seferihisar, Turkey [11], and a 4-point Likert scale is used as a metric for the visitors to answer (shown in Table 1). After careful inspection by specialists of the statements from Yurtseven et al., we decided to use these questions for this study with minor changes. To avoid answers without justification and to make clear the tourists' preference, we do not use the neutral (e.g., "no preference") answer, but instead we employ the 4-point scale to make a clearer distinction. For practical purposes, in order to ensure that the respondents have a clear understanding of the meaning of the statements, the statements were translated into Chinese, with some specific examples given to aid in the respondents' replies. For instance, for statement thirteen, "Promoting and Preserving local culture events," we cited a special event, Pas-ta'ay, given by the local Saisiyat tribe [2]. It is an incredibly special event that not many tourists have experienced, and it is a culture event that needs to be preserved. These statements translated to Chinese can be viewed in Table 1

The second part of the dataset contains information regarding the respondent's background, including gender, age, residence, education, duration of stay, how the visitor heard about the destination, their reason for visiting, and the amount of money spent during the visit. Each of these subjects contain two to six categorical answer options for statistical analysis.

The data was collected by my graduate student from Chung Hua University. The period of collection lasted from June 2020 to September 2020. Due to the Covid 19 pandemic, most of the activities were difficult to carry out. During the data collection period, the meaning of the statements and questions were explained if necessary, and the respondents were asked to answer based on their beliefs and experiences. In total, our dataset consists of 222 respondents.

2.2 Machine Learning Implemented in the Dataset

While the statistics software, SPSS, was an initial possibility for analyzing our data, we decided to use the Python programming language to apply the machine learning techniques for analysis to explore the usefulness of these techniques within the tourism field. Machine learning can be broken down into four primary categories: supervised learning, semi-supervised learning, unsupervised learning, and reinforcement learning. Due to the nature of this study, we rely on unsupervised learning algorithms. Quoted from [7]:

Unsupervised learning subsumes all kinds of machine learning where there is no known output, no teacher to instruct the learning algorithms. In unsupervised learning, the learning algorithm is just shown the input data and asked to extract knowledge from this data.

In this study, we used the k-means clustering algorithm, a hierarchical clustering algorithm, and Principal Component Analysis (PCA). The results of experimentation with these algorithms are given in later sections. Additionally, we applied descriptive statistics to the responses from the categorical questions (part two of the dataset). These responses were used to investigate whether the resulting groups from our experiments are also defined by similar backgrounds. The outcome is given in Section 3.

The primary purpose of this study was to make proper clusters for the visitors, grouping them based on their level of understanding of the concepts of slow cities or Slow Tourism using their response from the 15-statement, thereby gaining insight on the type of tourism that is occurring at this destination. From this, a proper marketing strategy for Slow Tourism to Nanzhuang can be designed. To accomplish this clustering, we utilized the k-means clustering algorithm preceded by a dimension reduction technique, Principal Component Analysis (PCA), and followed by the statistics application, one-way ANOVA. Quoted from [7]:

Clustering is the task of partitioning the dataset into groups, called clusters. The goal is to split up the data in such a way that points within a single cluster are very similar and points in different clusters are different. K-means clustering is one of the simplest and most commonly used clustering algorithms. It tries to find cluster centers that are representative of certain regions of the data.

Traditionally, machine learning is applied in engineering problems, but in recent years, artificial intelligence (AI) has become more popular in the social sciences. Due to this, more use of machine learning tools in these fields in addition to engineering is expected [6].

3. Main Results

3.1 Analysis of the Experimental Results

The dataset used in this study contains 222 data points. Of these 222 data points, 53.8% were female, and 46.2% were male. While the age group varied from under 18 years of age to over 50 years of age, the largest percentage were within the 23-30 age range (37.1%), followed by the 30-40 age range, and the 18-22 age range (29% and 19% respectively). In fact, 85% of all data points fell within the 18-40 age range. Most of the respondents in the data set come from either the Northern (43%), Central (40.3%), or Southern (14.5%) regions of Taiwan (97.8% of

the samples), with only 1.8% being from foreign countries, and only 0.4% being from some outlying island. Most of the respondents stayed in Nanzhuang for only a half-day (46.6%), with one day and two days, one night being the next most common duration (39.4% and 11.8% respectively). Online materials provided the widest exposure, with 37.6% of the respondents learning of Nanzhuang from online sources. This was followed closely by hearing from a friend (32.1%), with the remaining learning of the township because of proximity to their homes (16.7%), or from written materials (13.6%). Self-guided tours were the most common reason for travel (49.8%), followed by family trips and group tours (35.7% and 13.6% respectively). The largest separation occurred with the amount spent during the visit. The majority spent less than NTD 1,000 (65.2%), followed by 29.4% spending NTD 1,000-3,000. The remaining 5.4% spent NTD 3,000 or more. The graphical representations of all the descriptive statistics are given in Figure 1.

After calculating the initial descriptive statistics from the background information, the next task was our experimentation with clustering. We did not use scaling for the data since the answers are from the 4-point Likert scale, meaning there are no outliers in the data set. We used the Elbow Curve method to find the optimal number of clusters by calculating the sum of squared distances from the centroid for each number of k-clusters, see Figure 2. From this, it was determined that three clusters were best for representing our data points with k-means clustering. Another reason for not using too many clusters can be derived based on the explanation of the status of the tourists given in [10] for ecotourists, who also grouped tourists into three groups. Quoted from [10]:

“Harder” ecotourists reflect a high level of environmental commitment and affinities with wilderness type experiences, while “softer” ecotourists are much less committed on either dimension. “Structured” ecotourists, by comparison, reveal a strong pattern of commitment but a level of desire for interpretation, escorted tours, and services/ facilities that is usually more associated with mass tourism.

Due to a large dimension (i.e., 222*15) of the dataset, it is difficult to see the outcomes in a two-dimensional plane. To overcome this, we conducted Principal Component Analysis (PCA) to reduce the dimensionality of our data for visualization and two principal vectors are eventually selected. We chose the first two principal components for modeling, whose total explanation of variance was roughly 68.1% (61.4% for the first principal component, 6.7% for the second principal component). A scatter plot representation of the results of k-means clustering after PCA shows that the data points are well separated into three clusters on the two-dimensional plane, see Figure 4. Note that, the number of points is not the same as given by the numerical outputs of the number of respondents in each of the three groups due to the fact that these points are PCA results and only show 68.1% of the original data. Three cluster groups were obtained and the number of people in those three groups are given in Figure 5. We name these three groups as: *Advocates of Slow Tourism (AST)*; *Conscious of Slow Tourism (CST)* and *Unconscious of Slow Tourism (UST)*, based on the average scores of the three groups. The overall average scores of the three groups are 2.61, 2.60, and 2.46, respectively; and the number of tourists in the three groups are 95, 104, and 23, respectively. Though, the average scores of the three groups do not show a strong difference, in each sub-problem, as shown in Table 1, we can see that the average scores of the three groups do show a difference. In order to be more confident about the number

of groups given by the k-means algorithm, we also applied the one-way ANOVA (analysis of variance) to see if there is any difference in the means between the three groups in each sub-problem of the questionnaire. The results are summarized in Table 1. We can see that all the p-values in each question are small which confirms the rejection of the null hypothesis that the means are equal between the three groups and confirms that our assumption from the Elbow method is suitable for this dataset. This is further proof that the separation into three groups for the tourists is reasonable.

After properly separating the tourists into three groups, we further study the descriptive statistics of the tourists' background and related behavior in each of the three groups, to determine if they reveal any differences between the average scores similar to the performance of the different mean scores in the questionnaire among the three groups. For example, does the AST group have a higher educational background than the CST group or the UST group? Is the average income in the AST group higher than that of the UST or CST groups? In Table 2, we have provided both the median and the average in each of the three groups. The overall scores in either the median or average scores for the three groups do not show a strong difference, e.g., the ages of the AST group are not higher than the CST or UST groups. The median for all groups is 3.0, and the averages are 3.32, 3.28, and 3.13 for the AST, CST, and UST groups, respectively. For the level of education, the median is again 3.0 for all groups, and the averages are 2.74, 2.93, and 2.99 for the AST, CST, and UST groups, respectively. Interestingly, the mean score for educational background of the AST group is lower than the CST group. This shows that the educational background for the UST group is similar to the AST group, which is an interesting outcome. That means there are tourists who have a higher education but are still not aware of this fresh style of tourism. This implies that educational background is not a huge factor in one's appreciation of Slow Tourism and does not need extra consideration in a targeted marketing scheme. From these descriptive statistics we discovered that backgrounds of the three groups are not significantly different.

The final observation from the results begs the question, why are the numbers for both the AST and CST groups (they are 89.6% of all the tourists) relatively high? We believe it is related to the description of the questions. It is possible that the questions provide some positive images for tourists, and even though tourists may not actually know if these concepts are related to Slow Tourism, they may still favor that which is implied by the questions. That is why the UST has the fewest tourists, this outcome may not reflect the actual situation of the tourists regarding Slow Tourism, since Slow Tourism is still a new concept for ordinary tourists. If possible, in future studies, the statements in the questionnaire could be stated with positive and negative statements allowing for the respondents to think more carefully. This may provide a closer reflection of the correct percentages of each group.

4. Conclusion and Recommendation

Slow Tourism is a new practice in the tourism industry but is gaining popularity among tourists looking for a more relaxed and immersive experience. This practice can be related to the concept of a slow city and the principles of the Cittaslow organization.

This study used data that was collected from visitors to Nanzhuang township, Taiwan to seek more insight about the correct types of tourists visiting to this slow city nowadays, and to investigate how their motivation to visit Nanzhuang might be related to the principles of the Cittaslow. Overall, our dataset contained 222 data points, representing 222 visitors to Nanzhuang

township. We were able to divide these visitors into three groups based on each respondent's responses to a 4-point Likert scale questionnaire by utilizing both the k-means clustering algorithm and PCA for better visualization purposes. One-way ANOVA was further used to make justification of the outcomes of the cluster groups by studying the p-values in each subproblem, and the results additionally confirm that the use of three clusters is the correct design. One interesting finding related to the background of the respondents was that they do not show a strong difference. The findings indicates that the backgrounds between the three groups make little to no difference.

The top two scoring groups consisted of a significantly sizable percentage of the dataset, indicating that the Cittaslow principles do influence motivation for visiting this destination, and that a substantial portion can be considered slow tourists. Though more studies should be given to ensure the real situation of Slow Tourism in Taiwan nowadays. Nevertheless, outcomes of our study reveal that the separation of the tourists into three groups is reasonable and can be further used for the design of suitable marketing strategies to attract more tourists to participate in Slow Tourism. This in turn will allow the slow cities in Taiwan to gain more economic growth through a proper and effective marketing strategy.

Further research using similar methods of the distinct types of tourists to the other three slow cities of Taiwan (Dalin township, Fonglin township, and Sanyi township) could provide interesting comparisons to our findings in this article, as well as give insight into the practice of Slow Tourism in Taiwan as a whole.

Acknowledgements The authors thank Mr. Spencer Perkins for the Python programming and my graduate student Ms. S.I. Yang for collecting all the data.

References

- [1] Cittaslow (1999). *Cittaslow International Charter*. Orvieto-Italy: International Cittaslow.
- [2] Cittaslow Nanzhuang (2016). Retrieved from <https://www.cittaslow.org/network/nanzhuang>.
- [3] Hakka Affairs Council (2018). Retrieved from <https://english.hakka.gov.tw>
- [4] Heitmann, S., Robinson, P. and Povey, G. (2011). *Slow Food, Slow Cities and Slow Tourism*, edited in *Research Themes for Tourism*, CABI North American, Cambridge, MA 02139, USA Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.
- [5] Matos, R. (2004). Can slow tourism bring new life to alpine regions. *The tourism and leisure industry: Shaping the future*, 93-103.
- [6] Mor, M., Dalyot, S., and Ram, Y. (2023). Who is a tourist? Classifying international urban tourists using machine learning. *Tourism Management*, 95, 104689.
- [7] Müller, A.C. and Guido, S. (2016). *Introduction to Machine Learning with Python*. Sebastopol, CA: O'Reilly
- [8] Slow Food Foundation (2022). Retrieved from <https://www.slowfood.com/>
- [9] Slow Food Foundation. *Slow Food Manifesto* (1989). Retrieved from <https://www.slowfood.com/about-us/key-documents/>
- [10] Weaver, D. B. and Lawton, L. J. (2002). Overnight Ecotourist Market Segmentation in the Gold Coast Hinterland of Australia, *Journal of Travel Research*, Vol. 40, February 2002, 270-280

[11] Yurtseven, H.R. and Kaya, Ozan (2011). Slow Tourists: A Comparative Research Based on Cittaslow Principles, *American International Journal of Contemporary Research*, Vol. 1, No. 2; Sep. 2011. pp. 91-98

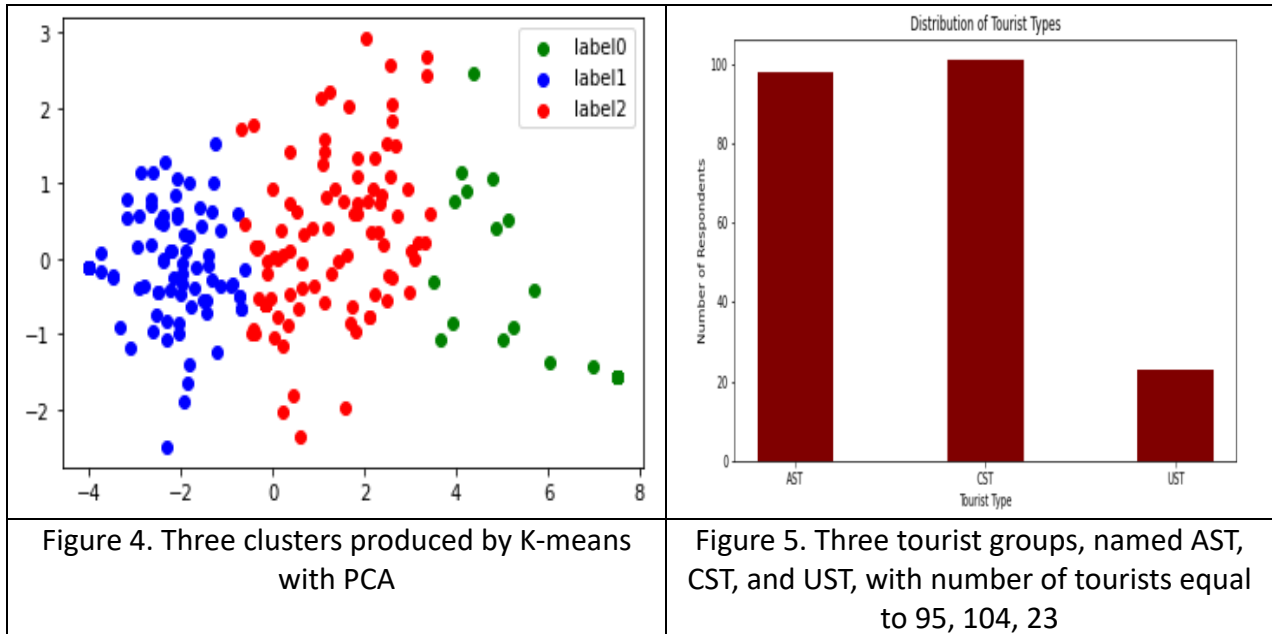
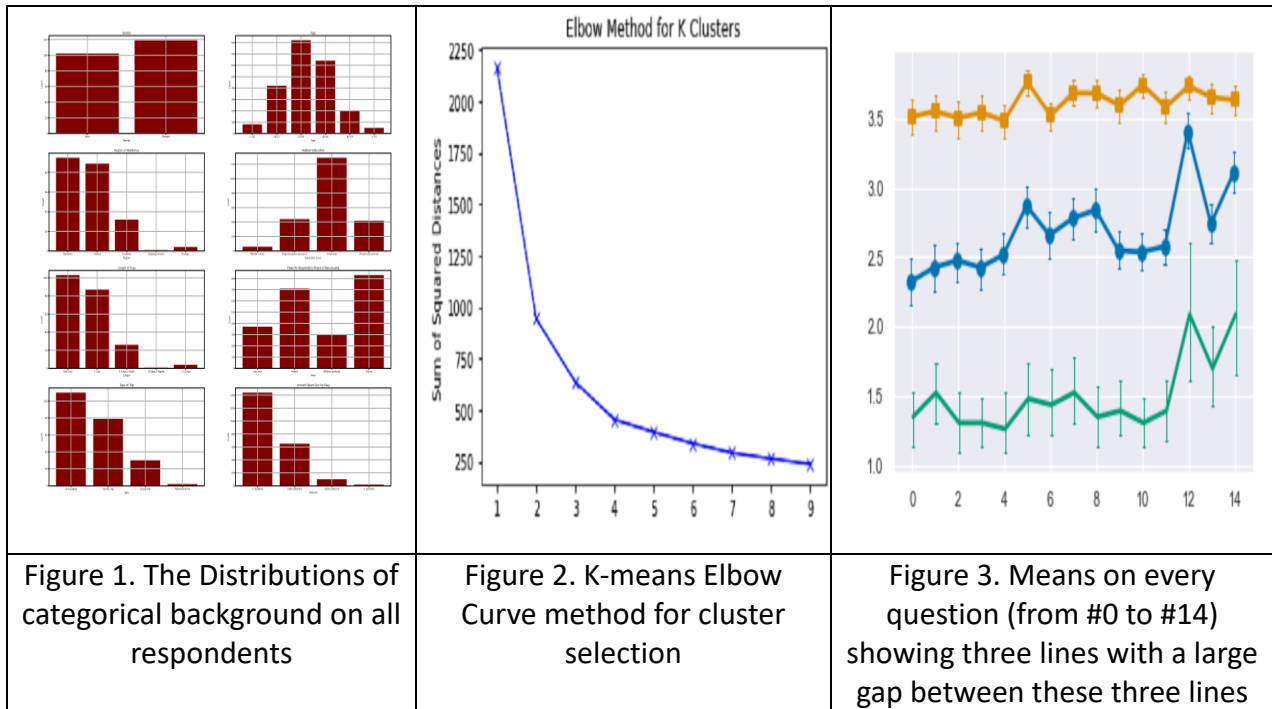


Table 1: Modified Cittaslow principles (adapted from [11])

	Mean AST	Mean CST	Mean UST	ANOVA F value	ANOVA p value
Ban on the use of O.G.M. in	3.51	2.32	1.35	116.36	3.71e-35

agriculture					
Plans for improving and for the reclamation of historical centers and/or works of cultural or historical value	3.55	2.42	1.52	106.35	5.31e-33
Quality green areas and service infrastructures (green areas, playgrounds, etc.)	3.50	2.47	1.30	107.86	2.48e-33
Plan for the distribution of merchandise and the creation of commercial centers for natural products	3.54	2.42	1.30	124.25	8.62e-37
Agreement with the shopkeepers regarding the reception and assistance to citizens in trouble: friendly shops	3.48	2.52	1.26	116.52	3.43e-35
Plans for the development of organic farming	3.76	2.87	1.48	143.58	1.43e-40
Certification of the quality of artisan produced products and objects and artistic crafts (Taiwanese Indigenous peoples' wood sculpture/knit textiles)	3.52	2.66	1.43	103.60	2.16e-32
Programs for the safeguarding of artisan and/or artistic craft products danger of extinction	3.68	2.78	1.52	130.37	5.09e-38
Safeguarding traditional methods of work and professions at a risk of extinction	3.68	2.84	1.35	134.13	9.29e-39
Use of organic products and/or those produced in the territory and the preservation of local traditions in restaurants, protected structures, and species and preparations risking extinction	3.59	2.55	1.39	149.70	1.04e-41
Favoring the activities of gastronomic Slow Food Presidia for species and preparations risking extinction	3.74	2.54	1.30	239.71	7.03e-56
Census of the typical products of the territory and support of their commercialization (updating of	3.58	2.58	1.39	152.97	2.66e-42

markets for local products, creation of appropriate spaces, e.g., farmer's market)					
Promoting and preserving local cultural events (festival: Pas-ta'ay event by Saisiyat tribe)	3.73	3.40	2.09	54.54	5.98e-20
Training courses for tourist information and quality hospitality, such as training for the local government staffs or having tourist center	3.65	2.74	1.70	103.66	2.09e-32
Preparation of slow itineraries of the city (brochures, websites, home pages, etc.)	3.63	3.11	2.09	45.58	3.45e-18
Overall	95	104	23		
N (total)	2.61	2.60	2.46		

Table 2. The median and average scores for tourists within three groups

	Median							
Item	Sex	Age	Residence	Education	Duration	How	Style	Amount
AST	2.0	3.0	4.0	3.0	2.0	2.0	4.0	1.0
CST	1.5	3.0	4.0	3.0	2.0	3.0	3.0	1.0
UST	1.0	3.0	4.0	3.0	1.0	2.0	3.0	1.0
	Mean (average)							
AST	1.60	3.32	4.23	2.93	1.69	2.31	3.48	1.36
CST	1.50	3.28	4.20	2.99	1.76	2.31	3.27	1.49
UST	1.48	3.13	4.30	2.74	1.57	2.09	3.13	1.26