Tourism Management 95 (2023) 104685

Contents lists available at ScienceDirect

Tourism Management

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

Optimal targeting of latent tourism demand segments

José M. Cazorla-Artiles [a,*], Juan L. Eugenio-Martin [b]

[a] Universidad de Las Palmas de Gran Canaria, Departamento de Análisis Económico Aplicado. C/ SauloTorón S/n, Fac. de Economía, Empresa y Turismo, Despacho D. 3.15, CP 35017 Las Palmas de Gran Canaria, Spain

[b] Universidad de Las Palmas de Gran Canaria, Departamento de Análisis Económico Aplicado. C/ SauloTorón S/n, Fac. de Economía, Empresa y Turismo, Despacho D. 2.06, CP 35017 Las Palmas de Gran Canaria, Spain

A R T I C L E I N F O

Keywords:
Latent demand
Target markets
Marketing campaign
Market segmentation
Tourism expenditure

A B S T R A C T

This paper identifies for the first time the optimal target markets employing the latent tourism demand expenditure, a novel concept in tourism literature. The study quantifies latent tourism demand between each pair of origin-destination through distinguishing by type of tourism and seasonality. It works with market shares that are estimated via a fractional regression model. Moreover, latent demand is clustered using a market segmentation approach based on a latent class regression. Finally, the optimal target markets are chosen depending on the expected latent tourism expenditure. The result has clear policy implications in terms of which markets are promoted, the optimal channels of communication and the maximum budget for each marketing campaign.

Funding source

We acknowledge the research grant ECO2016-80268-R funded by Ministerio de Economia y Competitividad in Spain..

1. Introduction

Tourism policymakers are continuously looking for new markets with growth potential. Traditionally their focus has been on total expenditure, because it produces a positive effect on employment and GDP at tourism destination. Total tourism expenditure can be calculated by multiplying total number of arrivals, daily tourist expenditure and length of stay. Each of these three factors are key elements in tourism policy strategies. Some destinations focus their efforts on attracting more tourists, whereas others are interested in promoting a specific tourist profile with a greater daily expenditure or length of stay.

This paper focuses on quantifying tourism potential growth. It is based on the shares method for revealing latent tourism demand (Eugenio-Martin & Cazorla-Artiles, 2020). Such methodology identifies and quantifies the latent tourism demand by origins. However, not all market origins may be equally interesting to be targeted. Some market origins may spend less days or less money. Ideally, the latent tourism demand figure should be combined with the expected tourism expenditure to reveal the optimal segments to be targeted. This is developed in this paper.

The challenge is to provide an aggregate result that it is based on individual information. Every tourist is unique, in terms of booking, party size, accommodation chosen, attractions visited, meals taken, or any other preferences on the way of spending her time at the destination may be different. However, decisionmakers need to choose target markets based on aggregate figures. Thus, moving from the individual level to a clustered set of tourists is required. Moreover, this paper also considers the seasonality and it shows that the latent demand varies dramatically by season, so that the optimal targeting should proceed accordingly.

The methodology is applied in the Canary Islands. The origin market chosen for the segmentation analysis is Madrid for sun and beach purposes, since it reports the highest potential after the application of the shares method. The paper ends with an illustration showing the potential growth for the clusters by season, according to accommodation type and their respective values in terms of arrivals, daily expenditure, length of stay and resulting total tourism expenditure.

The results are useful for policymakers because they quantify the expected return in terms of tourism expenditure for each origin, kind of tourism product and season. This expected return may be used to specify the maximum budget allocated for a tourism campaign or route subsidy for each tourism market.

* Corresponding author.
E-mail addresses: josemanuel.cazorla@ulpgc.es (J.M. Cazorla-Artiles), juanluis.eugenio@ulpgc.es (J.L. Eugenio-Martin).

https://doi.org/10.1016/j.tourman.2022.104685
Received 9 May 2022; Received in revised form 21 October 2022; Accepted 28 October 2022
Available online 9 November 2022

2. Literature review

The two main questions in marketing are demand volume (Song et al., 2019) and market segmentation (Dolnicar, 2002). The former has received considerable attention in the development of tourism demand models. In particular, tourism demand modelling has focused on the determinants of destination choice (Morley et al., 2014). From an individual perspective the random utility models are an alternative (Nicolau & Más, 2005), but from an aggregate perspective panel data and time series are usually employed (Li et al., 2017).

The concept of latent tourism demand is not new (Davies & Prentice, 1995), but the identification and quantification of the latent tourism market has been under-investigated in the literature. However, the shares method proposed by Eugenio-Martin and Cazorla-Artiles (2020) addresses this issue. This method estimates latent tourism demand as a shares method considering three steps: (i) participation rate share, (ii) preference shares by kind of tourism and (iii) the difference between the expected market share and the current market share at a destination. It identifies the size of latent tourism demand niches by origin market and kind of tourism.

The relevance of market segmentation research has increased considerably since the publication of benefit segmentation by Haley (1968). The market segmentation approach considers a heterogenous market as a composition of smaller homogeneous markets according to product preferences of the market segments (Smith, 1956). That is the reason why market research efforts have led to the creation of different clusters. Consequently, depending on demand patterns, a differentiated product for each cluster is identified that matches product characteristics with the demand preferences of the target group (Dolnicar, 2002).

There are several ways to proceed with market segmentation. But the two fundamental approaches are conceptual segmentation and data-driven, or post hoc, segmentation (Dolnicar, 2002). The conceptual approach is also known as a priori segmentation, in the field of market structure analysis (Myers & Tauber, 1977), which consists of grouping tourists based on a pre-established category, for instance: age, gender, income and education. The data-driven or post hoc approach requires the application of quantitative data analysis to the data available to define the tourism segments. This approach has received increasing academic attention.

Depending on the nature of the data available for market segmentation, the literature distinguishes between several segmentation criteria. The typical criteria for market segmentation can be summarized in four groups (Dolnicar, 2018): (i) geographic, which is a simple criterion where the consumer is assigned to a specific location. This assignment facilitates the marketing task because, for instance, communication messages can be segmented by regions; (ii) socio-demographic, which includes factors such as age, gender, income and education. In some industries, for instance, luxury goods (and its relation with a high income customer profile), this approach is very useful; (iii) psychographic, which consists of bundling consumers according to psychological criteria. In this regard, Haley (1985) focuses on the mind of the consumer, such as preferences, aspirations, interests, beliefs, and benefit (Haley, 1968), while Cahill (2006) studies the more complex area of lifestyle, which examines a combination of characteristics, such as travel motivations or perceived risk; and (iv) behavioral, which consists in analyzing the behaviour of actual customers. However, the main disadvantage of this approach is that it is based on actual or previous customers and not on potential ones. These criteria are usually applied in combination (Tkaczynski et al., 2009), hence it is important to distinguish the relevant variables of each criterion employed (Tkaczynski et al., 2010).

Another key issue in a data-driven segmentation process is choosing the algorithm technique, i.e. the determination of the algorithm, variables and number of clusters. As Dolnicar, Grün, and Leisch (2018) argues, the algorithm method applied in data-driven segmentation usually relies on cluster analysis, and to choose the correct one, dataset and segment characteristics must be taken into account. The methods can be classified into three groups: (i) distance-based methods, in which the distance between observation characteristics plays a significant role (see for instance, Prayag, 2012). In this case, the distance measure must be defined, and the most common are: euclidean distance, Manhattan or absolute distance, asymmetric binary distance, hierarchical methods, partitioning methods (e.g., k-means or neural networks), and hybrid approaches; (ii) model-based methods, in which segmentation is based on two assumptions of market segmentation. First, each market has a certain size, and second, the consumers of a certain segment have segment-specific characteristics; (iii) algorithm with integrated variable selection, in which the two previous methods consider that all variables included in the algorithm contribute to the segmentation solution. However, in this group that assumption does not hold, so a pre-process method is applied to identify the variables that really contribute to the segmentation (e.g., bi-clustering, variable selection procedure for clustering binary data or factor-analysis).

Variable selection plays a role in market segmentation solutions, and researchers therefore need to define criterion to decide what is relevant. An alternative is choosing subjective criterion, but there are three common approaches to select the characteristics which produce a better solution: (i) the filter method, which considers the relevance of the variables themselves using univariate statistics (e.g., chi-square, correlation coefficient, variance threshold, etc.); (ii) the wrapper method, which consists in optimizing a measure of model performance through an algorithm that produces different combinations of variables (e.g., sequential feature selection algorithms, genetic algorithms, etc.); (iii) the embedded method, which is similar to the wrapper method because it consists in optimizing an objective function, but this function is defined by the researcher (e.g., LASSO or decision tree).

The last technical issue to define in the segmentation algorithm is the number of clusters. A subjective alternative is to leave the decision to the researcher, but it is common to define a criterion to optimize (see Dolnicar, 2018), such as: the normalized entropy criterion (Celeux & Soromenho, 1996), Bayesian information criterion (BIC) (Schwarz, 1978) or the Akaike information criterion (AIC) (Akaike, 1973). The difference between AIC and BIC lies in the 'penalty' for including new parameters, so that, for models with more than seven parameters the penalty is greater according to BIC, which prefers simpler models.

In the tourism sector, the market segmentation has got a long tradition (Dolnicar, 2020; Gray, 1982). It has been considered by hoteliers and tourism destination managers to identify the optimal market niches to target. Such optimality requires a proper definition in terms of criteria that depends on the purpose of the segmentation. Hoteliers may be interested on price differentiation (Namin et al., 2020) for optimal revenue management (Denizci and Shi, 2019; Xu et al., 2019), profiling profitable hotel customers (Dursun & Caber, 2016) with recency, frequency and monetary indicators (RFM analysis), or maximizing satisfaction and loyalty (Paulose & Shakeel, 2021).

The strategy at the destination level is different. Tourism destinations may be interested in understanding the market by distinguishing motivations (Bieger & Laesser, 2002; Ramires, Brandão, & Sousa, 2018), emotions (Bigné & Andreu, 2004), activities (Derek et al., 2019), loyalty (Stylidis et al. 2020) or benefit segmentation (Frochot, 2005). Moreover, they may be interested in maximizing the profitability by attracting high spending tourists (Mok & Iverson, 2000; Jang et al., 2002), who may also be related with certain niches, such as sustainable tourism (Moeller et al., 2011; Nickerson et al., 2016), or the kind of destination (Laesser & Crouch, 2006).

Most studies classify the tourists in segments by cluster analysis and provide a description of the characteristics of each segment. Additionally, other studies estimate the socioeconomic underpinnings behind such classification. For instance, Molera and Albaladejo (2007) employ logit multinomial modelling for this purpose. Instead of running two independent methods, a latent class regression may be considered. The latent class regression includes covariates for predicting class membership (Dayton & Macready, 1988; Hagenaars and McCutcheon, 2002); this is a "one-step" technique because the latent class model and the covariates' coefficients are estimated simultaneously.

More recently, the tourism market segmentation is being favoured by new sources of information that may be obtained from social media, mobile phones tracking, or other internet sources (Ring et al., 2014). It provides new opportunities to improve the current knowledge of this field (see for instance Kirilenko et al., 2019).

3. Methodology

3.1. Overview of the methodology

The methodology in this paper is carried out in four consecutive stages. The first stage identifies and quantifies the latent tourism demand, i.e. the markets which have got potential for growth. Such markets are distinguished by origin, kind of tourism, and seasonality. In a second stage, any market of interest is chosen, either based on the latent demand size (as revealed in stage 1) or based on any other concern of the policy-maker.

The methodology relies on tourist expenditure surveys to consider the length of stay and the daily expenditure of each tourist. However, the large heterogeneity found in this kind of surveys require some sort of homogeneous classification to work more efficiently, especially for marketing purposes. Stage 2 employs the behavioral and socio-demographic variables of the tourists in a latent class regression to identify the clusters and assign the tourists to each of them. This stage allows profiling the latent demand into segments. In stage 3, capacity adjustments are incorporated so that latent demand figures can be accommodated to the current supply. This is useful for short run policies, but stage 3 may be halted for long run policy analysis. Finally, in stage 4, the optimal target markets are revealed and chosen by combining the variables of interest.

3.2. Stage 1: Screening of the main latent tourism demand niches

The first step seeks to quantify latent tourism demand niches. For this purpose, the shares method defined by Eugenio-Martin and Cazorla-Artiles (2020) is applied, but with some adjustments. In this paper, the concept of latent demand considers a seasonal approach, and three kinds of tourism are studied: sun and beach; nature-based; and culture-based; and the definition of latent tourism demand is slightly different. Latent tourism demand is therefore defined by:

$$L_{odkt} = T_{ot} \cdot S_{1ot} \cdot S_{2okt} \cdot (E[S_{3odkt}] - S_{3odkt}), \forall odkt$$

where L denotes latent tourism demand; T, denotes the number of trips; S_1, denotes participation in outbound tourism; S_2, denotes the share of preferences by kind of tourism; S_{3odk}, denotes the market share of a destination by kind of tourism; o, denotes origin market; d, denotes destination; k, denotes kind of tourism; and t denotes the month.

The estimation of market shares, $E[S_{3odkt}]$, must consider that market shares lie between zero and one. OLS regression cannot guarantee predicted values will lie within such an interval, so that an alternative regression method needs to be employed. One alternative suggested by Ferrari and Cribari-Neto (2004) is Beta regression. Unfortunately, Beta regression assumes that values 0 or 1 must take probability zero. In this paper, zero market share is likely to happen among some pairs of origin-destination, especially, once it is decomposed by purpose of visit.

Papke & Wooldridge, 1996 suggests using Bernoulli quasi-likelihood methods as a solution to this problem. In this paper, a sequence of N independent observations $\{(x_i, y_i) : i = 1, \ldots, N\}$ where $0 \leq y_i \leq 1$ is assumed. A pooled cross-section time series of the monthly shares of all pairs of origin-destination by purpose of visit is considered. For all pairs, it is assumed that:

$$E(y_i|x_i) = G(x_i\beta)$$

where $G(\cdot)$ is a known function satisfying $0 < G(z) < 1$ for all $z \varepsilon \mathbb{R}$. In our case, $G(z) \equiv \Phi(z)$, where $\Phi(\cdot)$ is the standard normal cumulative density function. The quasi-likelihood estimator of β is obtained from the following maximization problem:

$$\max_{\beta} \sum_{i=1}^{N} \{(1 - y_i)\log[1 - \Phi(x_i\beta)] + y_i \log[\Phi(x_i\beta)]\}$$

According to Papke & Wooldridge, 1996, such estimator is consistent for β provided. Moreover, it is asymptotically normal regardless of the distribution of $(y_i|x_i)$.

3.3. Stage 2: Clustering of latent demand tourists

The market segmentation procedure must consider objective variables that facilitate policymakers being able to define their clients (Pulido-Fernández & Sánchez-Rivero, 2010) and establish an appropriate marketing campaign, defining for example the region, the channel, and the kind of tourism products to offer. A data-driven segmentation approach is employed using a latent class regression model. To aid tourism communication policies, the segmentation variables considered are related to types of accommodation and the booking process; and the socio-demographic tourist profile is also considered.

Segmentation is undertaken through a latent class analysis (Lazarsfeld, 1950) and a latent class regression (Dayton & Macready, 1988; Hagenaars and McCutcheon, 2002). The estimation is conducted by the R library poLCA (version 1.4.1); to understand the estimation procedure in detail, see Linzer and Lewis (2011). Latent class analysis is a segmentation method for multivariate categorical data, where from a group of observed or statistically independent manifest variables the researcher can bundle the observations with similar responses. Latent class analysis assumes that the probability for the individual i from the cluster r has a set of outcomes J for the manifest variables is:

$$f(Y_i; \pi_r) = \prod_{j=1}^{J} \prod_{k=1}^{K_j} (\pi_{jrk})^{Y_{ijk}}$$

where K_j are the possible outcomes for j manifest variable; π_{jrk} is the class-conditional probability; Y_{ijk} takes value 1 if the individual i has outcome k to the manifest variable j and 0 otherwise; $i \in \{1, \ldots, N\}$; $j \in \{1, \ldots, J\}$; $k \in \{1, \ldots, K_j\}$; and $r \in \{1, \ldots, R\}$.

The probability density function (pdf) for individual i across all classes is calculated as the weighted sum:

$$P(Y_i|\pi, p) = \sum_{r=1}^{R} p_r \prod_{j=1}^{J} \prod_{k=1}^{K_j} (\pi_{jrk})^{Y_{ijk}}$$

where p_r is the probability of membership to latent class r, so that, $\sum_{r=1}^{R} p_r = 1$.

The log-likelihood function for the pdf indicated is:

$$\ln L = \sum_{i=1}^{N} \ln \sum_{r=1}^{R} p_r \prod_{j=1}^{J} \prod_{k=1}^{K_j} (\pi_{jrk})^{Y_{ijk}}$$

The latent class model estimates $\widehat{p}_r$ and $\widehat{\pi}_{jrk}$ by the maximization of the log-likelihood using expectation-maximization (EM) algorithm (Dempster et al., 1977). Once the outcome and class probabilities are estimated, Bayes' theorem is applied to calculate the probability that individual i belongs to each class:

$$\widehat{P}(r_i|Y_i) = \frac{\widehat{p}_r f(Y_i; \widehat{\pi}_r)}{\sum_{q=1}^{R} \widehat{p}_q f(Y_i; \widehat{\pi}_q)}$$

To run this model, it is necessary to establish the manifest variables and the number of clusters. Both decisions are taken by the researcher. In order to facilitate tourism policy, the manifest variables considered are the type of accommodation and the booking process, as well as four clusters. To understand the determinants of belonging to each cluster, a latent class regression is considered, which is a "one-step" technique because the clusters and the determinants are estimated simultaneously. This is relevant to the policymaker because it reveals the main characteristics that 'pull' the most relevant segment.

In latent class regression the p_r varies according to the individual's covariates. A latent class is defined as the base and the other latent classes are compared with respect to the former, assuming that log-odd ratios are a linear function of the covariates. Then:

$$p_{ri} = p_r(X_i;\beta) = \frac{e^{X_i\beta_r}}{\sum_{q=1}^{R} e^{X_i\beta_q}}$$

where, β_r denotes the vector coefficient for latent class r, whose reference class is $\beta_1 = 0$; and X_i is the vector covariates for individual i.

The log-likelihood function for the pdf indicated is similar to a latent class model but using $p_r(X_i;\widehat{\beta})$ instead of p_r:

$$\ln L = \sum_{i=1}^{N} \ln \sum_{r=1}^{R} p_r(X_i;\widehat{\beta}) \prod_{j=1}^{J} \prod_{k=1}^{K_j} (\pi_{jrk})^{Y_{ijk}}$$

The latent class regression estimates $\widehat{\beta}_r$ and $\widehat{\pi}_{jrk}$ by the maximization of the log-likelihood using a modified EM algorithm with a Newton-Raphson step, as establishes Bandeen-Roche et al. (1997).

Similar to a latent class model, in latent class regression, the posterior probabilities of class membership are obtained using Bayes' Theorem, but replacing p_r with the function $p_r(X_i;\widehat{\beta})$:

$$\widehat{P}(r_i|X_i;Y_i) = \frac{p_r(X_i;\widehat{\beta}) f(Y_i;\widehat{\pi}_r)}{\sum_{q=1}^{R} p_q(X_i;\widehat{\beta}) f(Y_i;\widehat{\pi}_q)}$$

3.4. Stage 3: Adjusted latent demand by capacity constraints

The four clusters for each kind of tourism must be adjusted according to capacity constraints between the origin-destination pair. These constraints are revealed when the tourism relation between the origin and destination are conditioned by the available accommodation. This happens when the tourist accommodation type is 'own accommodation', or the property belongs to friends or relatives.

Hence, latent tourism demand is adjusted as follows:

$$L_{odktc}\begin{cases} L_{odktc} & if\ c = \{Hotel, P2P\} \\ L_{odktc}\cdot\alpha_c & if\ c = \{Owner, VFR\} \end{cases}$$

where c denotes the cluster, and α_c denotes the capacity restriction. The capacity restriction is defined as the share for the tourists from o to d with own or friends and relatives' property accommodation at destination d, divided by the tourists from o that enjoy tourism according to the same type of accommodation.

3.5. Stage 4: Criteria for optimal targeting market segments

Finally, the methodology shows the results of each segment in terms of expenditure and length of stay. So that, policymakers can choose the optimal market segment by considering its tourism product.

4. Results

4.1. Data

This paper employs data from the resident tourism survey carried out by the Spanish Statistical Office (INE, 2020). The frequency of the survey is monthly, and it has been carried out since 2015; data from this research therefore spans 2015 to 2019, resulting in 347,035 observations. The respondents declare their tourism destination and their main reason for going on holiday, which is assumed to be related with the former. For the quantification of latent tourism demand, the microdata is weighed up using the corresponding elevation factor to obtain the aggregate number of trips at regional level (NUTS 2). This paper defines outbound tourism as national tourism made to a region different than the region of residence. A distance matrix has been built using the euclidean distance between the main cities of each region, and these are measured in thousands of kilometres. Also, the average temperature measured in Celsius degrees is considered at regional level using data from the State Meteorological Agency of Spain (AEMET).

4.2. Stage 1: Screening of the main latent tourism demand niches

The shares method for revealing latent tourism demand aims to define the potential market size and quantify the expected and current market share of a destination. In this paper, the starting point is to quantify the size of participation in outbound tourism demand by month.

In Table 1 three participation shares are defined: participation in domestic tourism ($S^{dom}_{1,ot}$), which includes people who enjoy tourism in their own region of residence; participation in outbound tourism ($S^{out}_{1,ot}$), which refers to people who travel to another region in Spain; and participation in international tourism ($S^{int}_{1,ot}$), which refers to journeys to another country. For simplicity, the seasonal effect considers four quarters: Pre-summer, which comprises the months of March, April and May; summer, which comprises June, July and August; Post-summer, for September, October and November; and winter for December, January and February.

Table 1 shows that summer is the period with the highest outbound (0.48) and international (0.11) tourism figures, on average. Madrid is the region with the highest values of outbound tourism for each season, but especially in summer with 0.82. The regions with the highest domestic tourism are Andalusia, the Canary Islands and Galicia.

Once the participation in tourism is understood, the shares method looks for the preferences by kind of tourism. These preferences are quantified with the motivation for going on holidays, and this paper studies the preferences for culture-based, nature-based and sun and beach tourism.

Table 2 shows the preferences shares by kind of tourism, which also includes visiting friends and relatives. Visiting friends and relatives (VFR) is the main purpose for going on holidays out of the region of residence, especially in winter (0.43). The second motivation for going on holidays is sun and beach tourism with marked seasonality in summer (0.25). Culture-based and Nature-based tourism are the least relevant and do not show marked seasonality.

Table 3 shows the market share for each region as destination, and distinguishes kind of tourism and season. Andalusia (Pre-summer 0.21) shows the highest values as a culture destination except for winter, when Madrid (0.24) is preferred. Castilla y Leon (Post-summer 0.26) is the favourite destination for nature-based tourism. Sun and beach tourism is led by Andalusia and Valencia, and the popularity of the Canary Islands in winter is also notable.

The expected S_3 value lies within 0 and 1, so that they may be estimated by a fractional regression model. Related to the estimation of S_3 the purpose of this study is to define a parsimonious model. Specifically, the general specification can be expressed as:

$$E[S_3odkt] = G(q_{dk}, d_{od}\cdot t, c_o)$$

where q_{dk} denotes the destination quality for enjoying k kind of tourism, also known as the alternative specific constant (ASC); d_{od} denotes the distance between origin and destination; t denotes the seasonal effect

Table 1
Shares of tourism participation by season (S_{1ot}) of NUTS 2 regions in Spain.

Origin	Pre-Summer			Summer			Post-Summer			Winter		
	$S^{dom}_{1,ot}$	$S^{out}_{1,ot}$	$S^{int}_{1,ot}$	$S^{dom}_{1,ot}$	$S^{out}_{1,ot}$	$S^{int}_{1,ot}$	$S^{dom}_{1,ot}$	$S^{out}_{1,ot}$	$S^{int}_{1,ot}$	$S^{dom}_{1,ot}$	$S^{out}_{1,ot}$	$S^{int}_{1,ot}$
Andalusia	0.73	0.18	0.08	0.77	0.16	0.07	0.68	0.21	0.11	0.74	0.18	0.08
Aragon	0.46	0.48	0.06	0.33	0.58	0.09	0.50	0.43	0.07	0.63	0.31	0.05
Asturias	0.50	0.46	0.04	0.40	0.52	0.08	0.43	0.50	0.07	0.59	0.35	0.05
Balearic Is.	0.44	0.39	0.17	0.52	0.33	0.15	0.47	0.39	0.14	0.39	0.44	0.17
Canary Is.	0.72	0.19	0.09	0.76	0.16	0.08	0.73	0.19	0.08	0.71	0.25	0.04
Cantabria	0.38	0.56	0.07	0.27	0.61	0.12	0.33	0.58	0.09	0.39	0.53	0.08
Castilla – LM.	0.36	0.56	0.07	0.22	0.71	0.06	0.31	0.64	0.05	0.38	0.54	0.08
Castilla Leon	0.47	0.47	0.06	0.41	0.51	0.08	0.52	0.41	0.07	0.57	0.38	0.05
Catalonia	0.62	0.23	0.15	0.58	0.24	0.18	0.64	0.22	0.14	0.61	0.21	0.18
Valencia	0.58	0.32	0.09	0.55	0.34	0.11	0.56	0.35	0.09	0.58	0.33	0.10
Extremadura	0.55	0.39	0.06	0.32	0.58	0.10	0.38	0.57	0.06	0.48	0.46	0.07
Galicia	0.66	0.25	0.09	0.69	0.18	0.13	0.69	0.23	0.08	0.71	0.23	0.06
Madrid	0.11	0.78	0.10	0.07	0.82	0.11	0.13	0.76	0.11	0.16	0.74	0.11
Murcia	0.32	0.60	0.08	0.40	0.51	0.09	0.35	0.59	0.06	0.30	0.61	0.08
Navarra	0.34	0.59	0.08	0.27	0.61	0.13	0.32	0.57	0.11	0.44	0.48	0.08
Basque Country	0.18	0.70	0.12	0.17	0.72	0.11	0.21	0.70	0.09	0.32	0.61	0.08
La Rioja	0.29	0.64	0.07	0.24	0.65	0.11	0.22	0.70	0.08	0.28	0.64	0.08
Average	0.45	0.46	0.09	0.41	0.48	0.11	0.44	0.47	0.09	0.49	0.43	0.08

Table 2
Shares of kind of tourism preferences by season (S_{2okt}) of NUTS 2 regions in Spain.

Origin	Pre-Summer				Summer				Post-Summer				Winter			
	Cul	Nat	S&B	VFR	Cul	Nat	S&B	VFR	Cul	Nat	S&B	VFR	Cul	Nat	S&B	VFR
Andalusia	0.15	0.02	0.02	0.40	0.14	0.08	0.15	0.25	0.14	0.03	0.08	0.39	0.18	0.02	0.02	0.36
Aragon	0.15	0.05	0.14	0.30	0.04	0.03	0.42	0.23	0.11	0.02	0.17	0.24	0.13	0.01	0.06	0.43
Asturias	0.12	0.04	0.05	0.27	0.09	0.06	0.20	0.26	0.11	0.05	0.11	0.32	0.08	0.03	0.02	0.47
Balearic Is.	0.08	0.03	0.02	0.48	0.06	0.08	0.02	0.50	0.14	0.03	0.03	0.53	0.18	0.01	0.05	0.48
Canary Is.	0.15	0.01	0.01	0.33	0.23	0.04	0.03	0.39	0.17	0.01	0.01	0.36	0.09	0.00	0.01	0.42
Cantabria	0.09	0.03	0.08	0.34	0.12	0.07	0.13	0.38	0.16	0.04	0.07	0.38	0.09	0.00	0.01	0.49
Castilla – LM.	0.12	0.03	0.08	0.32	0.05	0.04	0.46	0.21	0.09	0.03	0.16	0.37	0.11	0.04	0.03	0.46
Castilla Leon	0.12	0.04	0.04	0.43	0.08	0.05	0.34	0.30	0.13	0.05	0.20	0.33	0.14	0.02	0.01	0.45
Catalonia	0.10	0.04	0.04	0.30	0.13	0.08	0.19	0.28	0.10	0.08	0.09	0.36	0.13	0.03	0.02	0.46
Valencia	0.19	0.08	0.02	0.30	0.13	0.13	0.09	0.34	0.12	0.13	0.04	0.30	0.15	0.06	0.01	0.44
Extremadura	0.16	0.03	0.10	0.32	0.09	0.03	0.39	0.20	0.06	0.02	0.24	0.25	0.08	0.02	0.01	0.38
Galicia	0.16	0.09	0.05	0.29	0.10	0.05	0.24	0.28	0.10	0.04	0.09	0.29	0.09	0.04	0.02	0.34
Madrid	0.07	0.06	0.08	0.36	0.05	0.09	0.31	0.30	0.08	0.09	0.11	0.36	0.07	0.04	0.03	0.47
Murcia	0.11	0.04	0.14	0.24	0.09	0.08	0.27	0.24	0.12	0.08	0.09	0.26	0.13	0.05	0.05	0.34
Navarra	0.13	0.04	0.07	0.31	0.06	0.06	0.45	0.19	0.10	0.06	0.14	0.36	0.11	0.03	0.02	0.45
Basque Country	0.10	0.05	0.08	0.19	0.06	0.14	0.22	0.24	0.06	0.09	0.08	0.36	0.07	0.07	0.04	0.29
La Rioja	0.10	0.03	0.09	0.29	0.05	0.05	0.42	0.23	0.10	0.06	0.13	0.34	0.05	0.05	0.01	0.50
Average	0.12	0.04	0.06	0.32	0.09	0.07	0.25	0.28	0.11	0.05	0.11	0.34	0.11	0.03	0.02	0.43

over distance; and c_o denotes the role of climate at origin.

The results are shown in Table 4. The reference destination is Andalusia and with respect to it, the other destination's dummy shows the role of each ASC. The distance enters the model with a different effect by month because the last one is a multiplicative dummy. It shows the different role of distance by season and kind of tourism, in this sense, for sun and beach tourism, the distance has a marked negative effect from June to October. It means that in summer, when the climate is warmer, people prefer to travel near to where they live; whereas in winter, the opposite is true. Origin temperature has a negative sign, which supports the previous affirmation on distance, i.e. the better the climate at origin is, the less willingness there is to travel abroad. This is specified in a quadratic form so that as long as the temperature increases, its negative effect decreases non-linearly.

Table 5 provides an illustration of how this methodology quantifies latent tourism demand from Madrid to the Canary Islands for sun and beach purposes in 2019. Madrid shows growth potential as a tourism market, especially in the summer.

4.3. Stage 2: Clustering of latent demand tourists

Madrid is the market with the highest growth potential for sun and beach purposes. For this reason, the paper considers the segmentation of this market, so that a latent class regression model is applied on this region. The results are shown in Table 6 and Table 7. The manifest variables considered are the channel for booking accommodation and type of accommodation; the regressors are age, income and party size; and four clusters are considered.

Table 6 shows the latent class probabilities of latent class regression. The first class comprises 28% of total tourists and can be defined as 'Hotel', because the leading accommodation of this cluster is Hotel (86%), and the accommodation is booked through the hotel webpage, hotel phone and travel agency webpage. The second class comprises 32% and is defined as 'Owner', because the most relevant accommodation type is own property (47%), followed by VFR property (37%), and they do not participate in the booking process. The third cluster is P2P because their accommodation is a 'House P2P' (97%) and their booking process relies on P2P webpage (43%) and Owner's phone (31%). Finally, the fourth cluster is VFR because their accommodation is

Table 3
Current destination market share by kind of tourism and season (S_{3dkt}) of NUTS 2 regions in Spain.

Destination	Pre-Summer				Summer				Post-Summer				Winter			
	Cul	Nat	S&B	VFR	Cul	Nat	S&B	VFR	Cul	Nat	S&B	VFR	Cul	Nat	S&B	VFR
Andalusia	0.21	0.07	0.26	0.13	0.14	0.04	0.23	0.12	0.17	0.03	0.25	0.11	0.20	0.09	0.21	0.12
Aragon	0.05	0.15	0	0.06	0.05	0.16	0.00	0.03	0.08	0.16	0.00	0.05	0.05	0.26	0.00	0.05
Asturias	0.01	0.07	0.02	0.03	0.05	0.15	0.02	0.04	0.04	0.07	0.02	0.03	0.02	0.03	0.00	0.02
Balearic Is.	0.01	0.03	0.05	0.01	0.01	0.01	0.06	0.02	0.02	0.01	0.10	0.01	0.01	0.01	0.02	0.01
Canary Is.	0.01	0.01	0.08	0.01	0.02	0.01	0.04	0.01	0.01	0.02	0.09	0.01	0.02	0.03	0.27	0.01
Cantabria	0.05	0.10	0.04	0.03	0.04	0.08	0.06	0.03	0.04	0.08	0.08	0.03	0.02	0.02	0.03	0.03
Castilla – LM.	0.07	0.07	0.00	0.17	0.07	0.11	0.00	0.15	0.06	0.14	0.00	0.13	0.08	0.14	0.00	0.15
Castilla Leon	0.16	0.27	0.00	0.16	0.14	0.24	0.00	0.19	0.16	0.26	0.00	0.16	0.11	0.17	0.00	0.15
Catalonia	0.03	0.03	0.07	0.04	0.06	0.03	0.09	0.04	0.02	0.02	0.07	0.06	0.04	0.03	0.02	0.05
Valencia	0.06	0.02	0.41	0.06	0.04	0.02	0.41	0.08	0.05	0.02	0.34	0.08	0.04	0.04	0.40	0.07
Extremadura	0.06	0.06	0.00	0.05	0.06	0.04	0.00	0.06	0.04	0.03	0.00	0.05	0.04	0.06	0.00	0.03
Galicia	0.02	0.04	0.01	0.04	0.11	0.05	0.04	0.05	0.06	0.03	0.02	0.04	0.04	0.02	0.00	0.04
Madrid	0.13	0.00	0.00	0.13	0.11	0.00	0.00	0.08	0.16	0.01	0.00	0.14	0.24	0.04	0.00	0.17
Murcia	0.02	0.02	0.07	0.03	0.00	0.00	0.05	0.02	0.01	0.01	0.02	0.02	0.01	0.00	0.04	0.02
Navarra	0.03	0.04	0.00	0.02	0.02	0.02	0.00	0.02	0.01	0.09	0.00	0.04	0.03	0.03	0.00	0.02
Basque Country	0.06	0.01	0.00	0.04	0.06	0.02	0.01	0.03	0.05	0.02	0.01	0.04	0.06	0.02	0.01	0.04
La Rioja	0.02	0.01	0.00	0.02	0.01	0.01	0.00	0.02	0.01	0.01	0.00	0.02	0.01	0.02	0.00	0.02
Total	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

VFR property (99%), and they do not book.

Table 7 shows the relationship between covariates and clusters. The covariate effects must be interpreted in relative terms to the Hotel cluster. Regarding the owner cluster, older people are more likely to be included in this segment, but a larger party size has a negative effect. P2P is more likely in young people with lower income and when the trip comprises a larger party size. VFR is more common among young people and party size has also a positive influence in choosing VFR accommodation, with respect to a hotel.

Semi-parametric analysis of the post-estimation probabilities provide further knowledge of the characteristics of each segment. Fig. 1 shows such probabilities. It shows that the probability of latent class membership of “owner” grows with income and age, as expected. Moreover, the probability of choosing P2P accommodation grows with the party size, as also expected. The figure also shows that the probability of belonging to VFR decreases with income and age. Concerning P2P, it should be noted that the probability of choosing P2P decreases with age, but not with income. Finally, the probability of belonging to “Hotel” class grows with age and income up to a point where they decrease because they shift to the “owner” class. Moreover, as expected, the “Hotel” class probability decreases with the party size.

The individuals are grouped into the four different clusters, and their respective weights are applied to obtain the share of each cluster by seasonality. There are four seasons considered in the results: i) Pre-summer, for March, April and May; ii) Summer, for June, July and August; iii) Post-summer, for September, October and November; iv) and winter, for December, January and February. In Table 8 Madrid outbound tourism for sun and beach purposes is combined with latent tourism demand from Madrid to the Canary Islands for sun and beach purposes, so that the table shows the relevance of each cluster by season.

4.4. Stage 3: Adjusted latent demand by capacity constraints

The result shown in Table 8 must be adjusted to capacity constraints for the clusters Owner and VFR. Table 9 shows this adjustment, and both clusters lose relevance in terms of tourist arrivals.

4.5. Stage 4: Criteria for optimal targeting market segments

The last step of this methodology is choosing the optimal market. The criteria followed in this paper has been to choose the market with the highest latent tourism expenditure. For this purpose, the average values of daily expenditure by tourist and length of stay are calculated for each group. Total latent tourism expenditure is obtained by multiplying the daily expenditure times the length of stay and latent tourism demand. Table 10 shows latent tourism expenditure from Madrid to the Canary Islands for sun and beach purposes by cluster and season. According to this criterion, the optimal market in Pre-summer is Hotel; in summer it is P2P followed by Hotel; in Post-summer the Hotel is the optimal cluster again; and finally in winter Owner cluster has the highest total expenditure.

5. Conclusions

This study shows a new procedure to target markets based on the latent tourism expenditure. This is the first time that this kind of study is conducted. It extends the literature that was based on arrivals, whereas this study considers the expected latent expenditure as the criterion for targeting markets. This is useful when a destination is planning a marketing campaign or promoting or reinforcing new tourist routes.

Stage 1 of this methodology quantifies latent tourism demand by origin, kind of tourism and season. It allows policymakers to design their tourism product to promote in ‘valley seasons’ by minimizing seasonality effects for the destination. In stage 2 the previous markets are segmented into lower and more homogeneous ones. The variables chosen in the segmentation process are related to the booking process and the accommodation, which facilitates policy-making for communication channels. Stage 3 is an adjustment of the previous one because of capacity constraints, especially for those clusters that depend on having own-property accommodation or a friend relationship between origin and destination. Finally, in stage 4, the optimal target is chosen according to latent expenditure. If policymakers pursue the maximization of tourism expenditure at destination, this methodology indicates the optimal targets.

Further research is required, especially in relation to the identification of other criteria for choosing optimal markets. Moreover, the application in this paper concerns pre-Covid 19 tourists’ behaviour. It is interesting to test whether the latent demand, segmentation, or any other characteristic has been altered after Covid 19 took place.

Impact statement

This study shows a new procedure to target markets based on the latent tourism expenditure. This is the first time that this kind of study is conducted. It extends the literature that was based on arrivals, whereas this study considers the expected latent expenditure as the criterion for

Table 4
Estimates of the fractional response model.

Explicative Variables	Culture	Nature	Sun and Beach
Destination dummy			
Aragon	−1.0113*** [0.000]	0.0913 [0.405]	–
Asturias	−1.0945*** [0.000]	−0.0988 [0.357]	−1.6673*** [0.000]
Balearic Is.	−1.5559*** [0.000]	−1.1272*** [0.000]	−0.6588*** [0.000]
Canary Is.	−1.2314*** [0.000]	0.4774 [0.137]	0.0167 [0.916]
Cantabria	−1.1147*** [0.000]	−0.2771** [0.013]	−1.1599*** [0.000]
Castilla – LM.	−0.8575*** [0.000]	−0.4275*** [0.000]	–
Castilla Leon	−0.3365*** [0.000]	0.1582 [0.147]	–
Catalonia	−0.9535*** [0.000]	−0.4599*** [0.000]	−0.7129*** [0.000]
Valencia	−0.9758*** [0.000]	−0.7465*** [0.000]	0.2478*** [0.004]
Extremadura	−1.1121*** [0.000]	−0.4349*** [0.000]	–
Galicia	−0.8408*** [0.000]	−0.2530** [0.041]	−1.4776*** [0.000]
Madrid	−0.1602*** [0.010]	−0.2530*** [0.000]	–
Murcia	−1.8091*** [0.000]	−1.3685*** [0.000]	−1.0901*** [0.000]
Navarra	−1.4103*** [0.000]	−0.7257*** [0.000]	–
Basque Country	−0.8897*** [0.000]	−0.5840*** [0.000]	−2.0045*** [0.000]
La Rioja	−1.7113*** [0.000]	−1.3547*** [0.000]	–
Multiplicative dummy Month · Distance			
January	−0.3223** [0.042]	−0.9848*** [0.005]	−0.2946** [0.040]
February	−0.1944 [0.162]	−1.8871*** [0.000]	−0.2619 [0.159]
March	−0.3062* [0.056]	−1.1252*** [0.000]	−0.2929** [0.048]
April	−0.1465* [0.096]	−1.0765*** [0.000]	−0.1580 [0.277]
May	−0.0989 [0.211]	−1.0050*** [0.000]	−0.1650 [0.288]
June	−0.1677 [0.103]	−0.8537*** [0.001]	−0.3979*** [0.009]
July	−0.0019 [0.986]	−0.8261** [0.016]	−0.3137 [0.144]
August	−0.2087** [0.039]	−0.7820*** [0.007]	−0.5027** [0.019]
September	−0.1096 [0.185]	−0.7619** [0.019]	−0.4819*** [0.000]
October	−0.0559 [0.533]	−1.3835*** [0.000]	−0.363** [0.010]
November	−0.1829* [0.066]	−1.3086*** [0.000]	−0.1567 [0.354]
December	−0.2823** [0.027]	−1.7146*** [0.000]	−0.0611 [0.590]
Origin temperature	−0.0794*** [0.000]	−0.0826*** [0.000]	−0.0835*** [0.000]
Origin temperature 2	0.0022*** [0.000]	0.0022*** [0.000]	0.0027 [0.000]
Pseudo R^2	0.1185	0.1198	0.1837
Number of observations	3264	3200	1881

P-values are in square brackets.
*** Level of significance 1%.
** Level of significance 5%.
* Level of significance 10%.

Table 5
Latent tourism demand from Madrid to the Canary Islands for sun and beach purposes in 2019.

Month	T_t	S_{1t}	S_{2t}	$E[S_{3t}]$	S_{3t}	L_t
January	2,100,041	0.75	0.02	0.20	0.06	4296
February	2,037,128	0.75	0.03	0.18	0.39	−10,013
March	2,374,564	0.76	0.02	0.15	0.40	−7337
April	3,362,460	0.81	0.08	0.21	0.02	41,627
May	3,151,547	0.77	0.13	0.19	0.02	51,435
June	2,791,572	0.79	0.11	0.11	0.02	23,690
July	4,156,644	0.83	0.34	0.20	0.03	197,159
August	5,599,308	0.82	0.39	0.10	0.02	150,364
September	2,932,950	0.78	0.19	0.08	0.02	27,486
October	2,103,492	0.75	0.06	0.11	0.04	6432
November	2,031,597	0.74	0.03	0.23	0.17	2919
December	2,702,553	0.72	0.03	0.29	0.29	−124

Table 6
Latent class probabilities of latent class regression model.

Manifest variables	Hotel	Owner	P2P	VFR
Latent class probabilities	0.2863	0.3243	0.2125	0.1777
Book accommodation				
- P2P webpage	0.0000	0.0000	0.4361	0.0000
- Hotel webpage	0.3763	0.0000	0.0000	0.0000
- Hotel phone	0.2593	0.0000	0.0000	0.0000
- Owner's webpage	0.0000	0.0000	0.1053	0.0000
- Owner's phone	0.0000	0.0000	0.3159	0.0000
- Travel agency webpage	0.2680	0.0000	0.0000	0.0000
- Travel agency phone	0.0964	0.0000	0.0000	0.0000
- Real estate webpage	0.0000	0.0000	0.0489	0.0000
- Real estate phone	0.0000	0.0000	0.0904	0.0000
- Does not book	0.0000	1.0000	0.0030	1.0000
Accommodation				
- B&B	0.0016	0.0000	0.0000	0.0000
- Rural B&B	0.0292	0.0000	0.0096	0.0000
- Camping	0.0593	0.0000	0.0000	0.0014
- Cruise	0.0000	0.0033	0.0000	0.0000
- Room P2P	0.0000	0.0000	0.0181	0.0000
- Hotel	0.8609	0.1349	0.0000	0.0000
- Other market acc.	0.0087	0.0000	0.0000	0.0000
- Other non-market acc.	0.0000	0.0018	0.0000	0.0060
- Hostel	0.0403	0.0007	0.0000	0.0000
- House P2P	0.0000	0.0032	0.9723	0.0000
- Own property	0.0000	0.4721	0.0000	0.0000
- VFR property	0.0000	0.3795	0.0000	0.9922
- Shared property	0.0000	0.0045	0.0000	0.0004

Table 7
Latent class probabilities of latent class regression model.

Class	Covariates	Coefficient	Std. error	t Value	Pr(>|t|)
Owner/Hotel	(Intercept)	−0.01416	0.00518	−2.734	0.010
	Age	0.01565	0.00173	9.041	0.000
	Income	0.00005	0.00003	1.561	0.128
	Party size	−0.28906	0.03610	−8.008	0.000
P2P/Hotel	(Intercept)	0.01251	0.00369	3.390	0.002
	Age	−0.01286	0.00197	−6.532	0.000
	Income	−0.00007	0.00003	−2.202	0.035
	Party size	0.18018	0.03285	5.484	0.000
VFR/Hotel	(Intercept)	−0.00562	0.00265	−2.119	0.042
	Age	−0.02212	0.00340	−6.511	0.000
	Income	−0.00007	0.00004	−1.578	0.124
	Party size	0.21143	0.04250	4.975	0.000
Number of obs.			4419		
Number of estimated parameters			96		
Residual degrees of freedom			33		
Maximum log-likelihood			−10,336.49		
AIC			20,864.98		
BIC			21,478.77		
χ^2			95.47003		

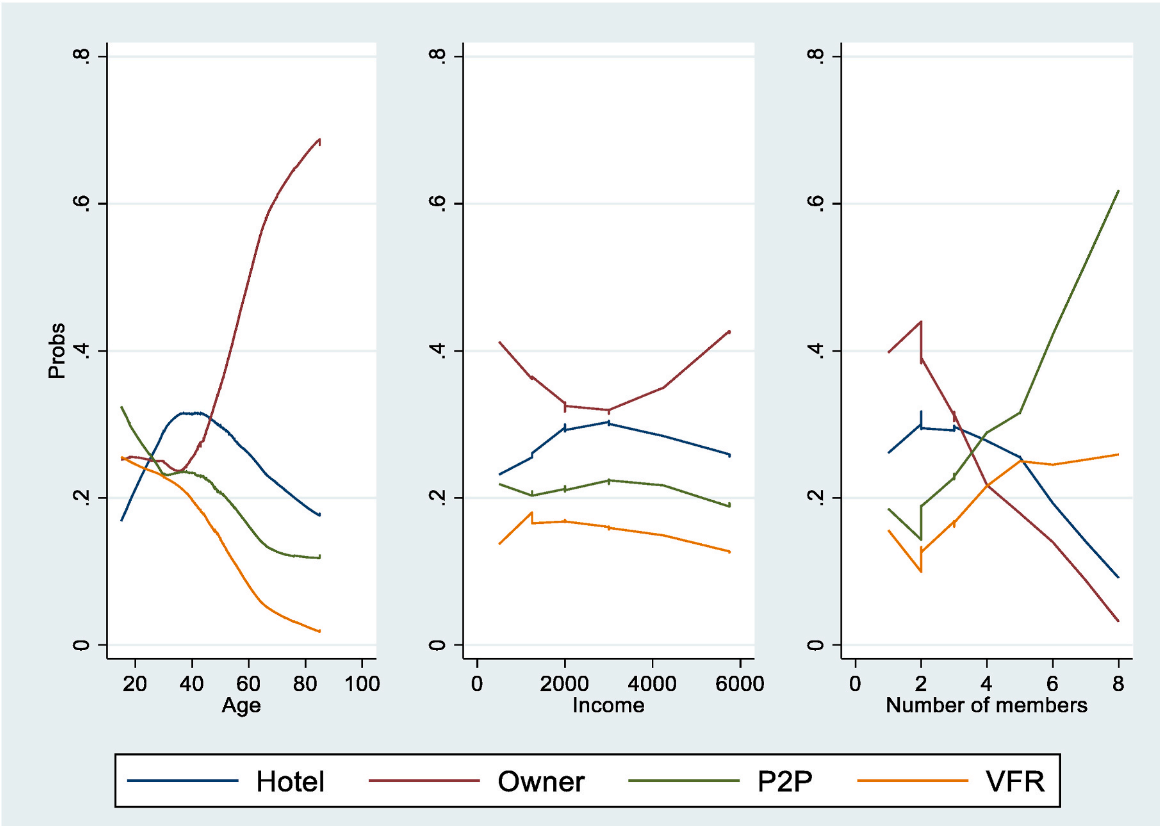

Fig. 1. Semi-parametric analysis of the probabilities of latent class membership by age, income and party size.

Table 8
Latent tourism demand from Madrid to the Canary Islands for sun and beach purposes by cluster and season.

Season	L_t	Hotel	Owner	P2P	VFR
Pre-summer	85,725	25,091	26,577	10,892	23,165
Summer	371,213	92,545	80,635	106,918	91,115
Post-summer	36,837	11,815	12,439	6803	5780
Winter	−5841	−2059	−2587	−387	−808

Table 9
Adjusted latent tourism demand from Madrid to the Canary Islands for sun and beach purposes by cluster and season.

Season	L_t	Hotel	Owner	P2P	VFR
Pre-summer	85,725	25,091	3466	10,892	14,121
Summer	371,213	92,545	10,516	106,918	55,542
Post-summer	36,837	11,815	1622	6803	3523
Winter	−5841	−2059	337	−387	−493

targeting markets. It distinguishes the kind of tourism, the origin region, the season and the market segment. This is useful when a destination is planning a marketing campaign or promoting or reinforcing new tourist routes.

For instance, it may estimate that the region of Madrid is an ideal target market to visit the Canary Islands for sun and beach purposes, for the segment of tourists who stay in hotels or P2P, but only during Summer. The latent demand in Winter is null and during pre-summer and post-summer periods is quite low. The rest of the market segments are less important. It is useful as a tool for tourism policymakers to optimize their marketing efforts.

Table 10
Latent tourism expenditure from Madrid to the Canary Islands for sun and beach purposes by cluster and season.

Pre-summer	Daily expenditure	Length of stay	L	Total expenditure
Hotel	98.70	4.12	25,091	10,203,105
Owner	51.90	5.44	3466	978,577
P2P	68.60	4.52	10,892	3,377,304
VFR	31.90	5.23	14,121	2,355,905
Summer	Daily expenditure	Length of stay	L	Total expenditure
Hotel	83.50	6.62	92,545	51,156,100
Owner	37.50	10.90	10,516	4,298,415
P2P	55.40	9.37	106,918	55,500,920
VFR	25.10	9.99	55,542	13,927,101
Post-summer	Daily expenditure	Length of stay	L	Total expenditure
Hotel	83.60	7.05	11,815	6,963,525
Owner	34.60	13.60	1622	763,248
P2P	58.00	9.34	6803	3,685,321
VFR	28.60	10.2	3523	1,027,730
Winter	Daily expenditure	Length of stay	L	Total expenditure
Hotel	113.00	4.51	−2059	−1,049,328
Owner	74.30	5.70	337	142,723
P2P	95.70	4.81	−387	−178,143
VFR	38.30	3.50	−493	−66,087

Statement of contribution

José Manuel Cazorla-Artiles contributed with conceptualization, writing, programming and data management. Juan L. Eugenio-Martin contributed with conceptualization and writing.

Declaration of competing interest

None.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. Petrov, & F. Csake (Eds.), *Second international symposium on information theory* (pp. 267–281). Budapest, Hungary: Akademiai Kiado.

Bandeen-Roche, K., Miglioretti, D. L., Zeger, S. L., & Rathouz, P. J. (1997). Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association, 92*(440), 1375–1386.

Bieger, T., & Laesser, C. (2002). Market segmentation by motivation: The case of Switzerland. *Journal of Travel Research, 41*, 68–76.

Bigné, J. E., & Andreu, L. (2004). Emotions in segmentation: An empirical study. *Annals of Tourism Research, 31*(3), 682–696.

Cahill, D. J. (2006). *Lifestyle market segmentation*. New York: Haworth Press.

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification, 13*(2), 195–212.

Davies, A., & Prentice, R. (1995). Conceptualizing the latent visitor to heritage attractions. *Tourism Management, 16*(7), 491–500.

Dayton, C. M., & Macready, G. B. (1988). Concomitant-variable latent-class models. *Journal of the American Statistical Association, 83*(401), 173–178.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via EM algorithm. *Journal of the Royal Statistical Society: Series B, 39*(1), 1–38.

Denizci Guillet, B., & Shi, X. (2019). Can revenue management be integrated with customer relationship management? *International Journal of Contemporary Hospitality Management, 31*(2), 978–997.

Derek, M., Woźniak, E., & Kulczyk, S. (2019). Clustering nature-based tourists by activity. Social, economic and spatial dimensions. *Tourism Management, 75*, 509–521.

Dolnicar, S. (2002). A review of data-driven market segmentation in tourism. *Journal of Travel & Tourism Marketing, 12*(1), 1–22.

Dolnicar, S. (2020). Market segmentation analysis in tourism: A perspective paper. *Tourism Review, 75*(1), 45–48.

Dolnicar, S., Grün, B., & Leisch, F. (2018). *Market segmentation analysis: Understanding it, doing it, and making it useful*. Singapore: Springer.

Dursun, A., & Caber, M. (2016). Using data mining techniques for profiling profitable hotel customers: An application of RFM analysis. *Tourism Management Perspectives, 18*, 153–160.

Eugenio-Martin, J. L., & Cazorla-Artiles, J. M. (2020). The shares method for revealing latent tourism demand. *Annals of Tourism Research, 84*, 1–17.

Ferrari, S., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics, 31*, 799–815.

Frochot, I. (2005). A benefit segmentation of tourists in rural areas: A scottish perspective. *Tourism Management, 26*(3), 335–346.

Gray, H. P. (1982). The contributions of economics to tourism. *Annals of Tourism Research, 9*(1), 105–125.

Hagenaars, J. A., & McCutcheon, A. L. (Eds.). (2002). *Applied latent class Analysis*. Cambridge: Cambridge University Press.

Haley, R. I. (1968). Benefit segmentation: A decision-oriented research tool. *Journal of Marketing, 32*(3), 30–35.

Haley, R. I. (1985). *Developing effective communications strategy – a benefit segmentation approach*. New York: Wiley.

INE. (2020). *Encuesta de turismo de residentes*. Spain: FAMILITUR. Instituto Nacional de Estadística. http://www.ine.es.

Jang, S. C., Morrison, A. M., & O'Leary, J. T. (2002). Benefit segmentation of Japanese pleasure travelers to the USA and Canada: Selecting target markets based on the profitability and risk of individual market segments. *Tourism Management, 23*(4), 367–387.

Kirilenko, A. P., Stepchenkova, S. O., & Hernandez, J. M. (2019). Comparative clustering of destination attractions for different origin markets with network and spatial analyses of online reviews. *Tourism Management, 72*, 400–410.

Laesser, C., & Crouch, G. I. (2006). Segmenting markets by travel expenditure patterns: The case of international visitors to Australia. *Journal of Travel Research, 44*(4), 397–406.

Lazarsfeld, P. F. (1950). The logical and mathematical foundations of latent structure analysis. In S. A. Stouffer (Ed.), *Measurement and prediction* (pp. 362–412). New York: John Wiley and Sons.

Linzer, D. A., & Lewis, J. B. (2011). poLCA: An R package for polytonomous variable latent class Analysis. *Journal of Statistical Software, 42*(10), 1–29.

Li, H., Song, H., & Li, L. (2017). A dynamic panel data analysis of climate and tourism demand: Additional evidence. *Journal of Travel Research, 56*(2), 158–171.

Moeller, T., Dolnicar, S., & Leisch, F. (2011). The sustainability-profitability trade-off in tourism: Can it be overcome? *Journal of Sustainable Tourism, 19*(2), 155–169.

Mok, C., & Iverson, T. J. (2000). Expenditure-based segmentation: Taiwanese tourists to Guam. *Tourism Management, 21*(3), 299–305.

Molera, L., & Albaladejo, I. P. (2007). Profiling segments of tourists in rural areas of South-Eastern Spain. *Tourism Management, 28*(3), 757–767.

Morley, C., Rosselló, J., & Santana-Gallego, M. (2014). Gravity models for tourism demand: Theory and use. *Annals of Tourism Research, 48*, 1–10.

Myers, J. H., & Tauber, E. (1977). *Market structure analysis*. American Marketing Association.

Namin, A., Gauri, D. K., & Kwortnik, R. J. (2020). Improving revenue performance with third-degree price discrimination in the cruise industry. *International Journal of Hospitality Management, 89*, Article 102597.

Nickerson, N. P., Jorgenson, J., & Boley, B. B. (2016). Are sustainable tourists a higher spending market? *Tourism Management, 54*, 170–177.

Nicolau, J. L., & Más, F. J. (2005). Stochastic modelling. A three-stage tourist choice process. *Annals of Tourism Research, 32*, 49–69.

Papke, L. E., & Wooldridge, J. M. (1996). Econometrics methods for fractional response variables with an application to 401(K) plan participation rates. *Journal of Applied Econometrics, 11*, 619–632.

Paulose, D., & Shakeel, A. (2021). Perceived experience, perceived value and customer satisfaction as antecedents to loyalty among hotel guests. *Journal of Quality Assurance in Hospitality & Tourism, 23*(2), 447–481.

Prayag, G. (2012). Paradise for who? Segmenting visitors' satisfaction with cognitive image and predicting behavioural loyalty. *International Journal of Tourism Research, 14*(1), 1–15.

Pulido-Fernández, J. I., & Sánchez-Rivero, M. (2010). Attitudes of the cultural tourist: A latent segmentation approach. *Journal of Cultural Economics, 34*(2), 111–129.

Ramires, A., Brandão, F., & Sousa, A. C. (2018). Motivation-based cluster analysis of international tourists visiting a World Heritage City: The case of Porto, Portugal. *Journal of Destination Marketing & Management, 8*, 49–60.

Ring, A., Tkaczynski, A., & Dolnicar, S. (2014). Word-of-Mouth segments: Online, offline, visual or verbal? *Journal of Travel Research, 55*(4), 481–492.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*(2), 461–464.

Smith, W. R. (1956). Product differentiation and market-segmentation as alternative marketing strategies. *Journal of Marketing, 21*(1), 3–8.

Song, H., Qiu, R. T., & Park, J. (2019). A review of research on tourism demand forecasting: Launching the Annals of Tourism Research Curated Collection on tourism demand forecasting. *Annals of Tourism Research, 75*, 338–362.

Stylidis, D., Woosnam, K. M., & Ivkov, M. (2020). Tourists' emotional solidarity with residents: A segmentation analysis and its links to destination image and loyalty. *Journal of Destination Marketing & Management, 17*, Article 100458.

Tkaczynski, A., Rundle-Thiele, S. R., & Beaumont, N. (2009). Segmentation: A tourism stakeholder view. *Tourism Management, 30*(2), 169–175.

Tkaczynski, A., Rundle-Thiele, S. R., & Beaumont, N. (2010). Destination segmentation: A recommended two-step approach. *Journal of Travel Research, 49*(2), 139–152.

Xu, X., Zhang, L., Baker, T., Harrington, T. J., & Marlowe, B. (2019). Drivers of degree of sophistication in hotel revenue management decision support systems. *International Journal of Hospitality Management, 79*, 123–139.

José Manuel Cazorla-Artiles PhD is Lecturer at the University of Las Palmas of Gran Canaria, Spain. His research interest is tourism economics. He has participated in international research projects for European Investment Bank or Inter-American Development Bank. He has published papers in leading journals, such as Annals of Tourism Research or Tourism Economics.

Juan Luis Eugenio-Martin PhD is Senior Lecturer in Tourism Economics at the University of Las Palmas of Gran Canaria, Spain. His research interest is tourism economics. He has carried out international research projects for The World Bank, European Commission, or Inter-American Development Bank. He has published papers in leading journals, such as Tourism Management, Annals of Tourism Research, Tourism Economics or Journal of Travel Research, among others.