ISSN (PRINT): 1994 - 7658 ISSN (ONLINE): 1314 - 0817

European Journal of Tourism Research
http://ejtr.vumk.eu

Competitiveness and overtourism: a proposal for an early warning system in Spanish urban destinations

José Francisco Perles-Ribes[1]*, Ana Belén Ramón-Rodríguez[2], Luis Moreno-Izquierdo[3] and María Jesús Such-Devesa[4]

[1] Department of Applied Economic Analysis, Faculty of Economics and Business Sciences, University of Alicante, E-mail: jose.perles@ua.es
[2] Department of Applied Economic Analysis, Faculty of Economics and Business Sciences, University of Alicante, E-mail: anar@ua.es
[3] Department of Applied Economic Analysis, Faculty of Economics and Business Sciences, University of Alicante, E-mail: luis.moreno@ua.es
[4] Department of Economics, Faculty of economics, business and tourism, University of Alcalá. E-mail: mjesus.such@uah.es

* Corresponding author

Abstract
The tourism industry is undergoing accelerated changes that pose significant challenges for both destination and business managers as well as for researchers of the tourism phenomenon. Two of these challenges that are particularly relevant are the emergence of the sharing economy and its influence on the degree of overtourism perceived in the tourist destinations. This paper addresses the subject through the use of machine learning techniques. The findings show that machine learning techniques are especially well-suited tools for dealing with these kinds of tourism issues. The findings also show that for the Spanish case, tourism competitiveness is a key predictor of overtourism

Keywords: Overtourism, Competitiveness, Early-warning system, Spain, Machine Learning

Citation: Perles-Ribes, J.F., Ramón-Rodríguez, A.B., Moreno-Izquierdo, L. and Such-Devesa. M.J. (2021). Competitiveness and overtourism: a proposal for an early warning system in Spanish urban destinations. *European Journal of Tourism Research* 27, 2707.

Introduction

The tourism industry is undergoing accelerated changes that pose significant challenges for both destination and business managers as well as researchers of the tourism phenomenon. Before the COVID-19 outbreak, specifically the emergence of the sharing economy and its influence on the degree of overtourism perceived in the tourist destinations are two of the most salient challenges (Huete and Mantecón, 2018; Koens, Postma and Papp, 2018; Milano, 2018; Oklevik *et al.*, 2019).

Although the current situation is very new and changing at a fast pace, it can be considered to be rooted in the race for competitiveness of destinations that began in the mid-1990s. The publication of Michael Porter's book (1990) on the competitiveness of territories and the adaptation of this model to tourist destinations by Crouch and Ritchie (1999) encouraged destination managers to compete for the success of their destinations.

However, when it came to implementation, in many places the essential aspects of long-term viability and sustainability advocated by these theories were neglected, and priority was given to short-term economic and profitability aspects, which were easier to measure and assess in terms of arrivals, income or market shares achieved (Hall, Gössling and Scott, 2015; Hunter, 2002). The consequence, in many places, was the implementation of a misunderstood competitiveness policy that has led to an exacerbated growth of the destinations without taking into account the fundamental element of the theories: that the purpose of this development was to improve the standard of living of the residents of the destinations.

Under these premises, in many destinations, especially those most highly developed in terms of their competitiveness, a negative sentiment towards the tourism phenomenon has arisen in local societies, which are increasingly questioning whether the benefits in terms of income or employment generated by tourism are sufficiently relevant to compensate for the negative effects that it causes in terms of environmental degradation or the saturation of the destinations (Sharpley, 2014).

This article explores the relationship between these two phenomena in urban tourist destinations in Spain, one of the most competitive tourist countries in the world (World Economic Forum, 2019). Spain's high degree of tourism competitiveness allows us to study, in depth, the effects of the emergence of collaborative economy platforms in certain cities on tourists and residents. Furthermore, the case of Spanish urban destinations is interesting insofar as almost all of them have reached a high degree of tourism competitiveness. Some destinations (e.g. Barcelona, Balearic Islands or to a lesser extent Valencia) have observed a distinct manifestation of overtourism while others (e.g. Madrid or Malaga) are more reluctant to consider that they might be suffering from overtourism. This fact justifies using them as a case study. Thus, it enables us to observe when and under what circumstances a high volume of tourists concentrated in a destination leads to problems of saturation, overtourism or even tourismophobia.

To date, the debate and research on overtourism has taken a qualitative perspective and has mainly been based on the comparison of different case studies (e.g. Milano, 2018; Alcalde-García *et al*, 2018; Peeters *et al*, 2018; UNWTO *et al*, 2018; Neuts and Nijkamp, 2012; Popp, 2012; Seraphin, Sheeran and Pilato, 2018). This paper addresses this phenomenon from a quantitative viewpoint.

The most complete study to date on overtourism is probably the work of Peeters *et al* (2018). Their extensive study for the European Parliament analyses practically all aspects related to this phenomenon, proposing a definition and a conceptual model on overtourism, detailing its main causes and consequences and proposing corrective measures for the problems generated by it. The study also addresses the measures to combat overtourism, for which a wide range of indicators is proposed which, following the initial efforts in this area, are analysed using descriptive statistical techniques. Finally, the study considers and compares a series of case studies.

The conceptual model of Peeters *et al* (2018) and their measures to address overtourism are key references for this article. However, this study focuses on two specific aspects of overtourism: the determinants or causes of overtourism through the proposal of a more simplified model in which competitiveness and the sharing economy play a key role; and the measurement of overtourism through the use of advanced statistical techniques for the analysis and prediction of the phenomenon. This paper also focuses on a more specific territorial area, the case of Spain's urban tourist destinations.

Specifically, the aim of the paper is twofold. First, following the suggestion of Peeters *et al.* (2018) the article attempts to determine the extent to which overtourism could be interpreted as a synonym for an excess of competitiveness or an excess of success of a destination when this competitiveness does not translate into an improvement of the living conditions of the resident population or when the diseconomies generated by the mass influx of visitors do not compensate the benefits generated by tourism or offer a positive experience to the tourist. Therefore, the research questions of this study are:

RQ1 "Is overtourism a consequence of an excess of success in competitiveness of the destination?"
RQ2 "Is the sharing economy and its development in a destination a cause of boredom of the population?"

The questions are related to the sharing economy phenomenon that has emerged in tourism and has generated new relationships between the agents involved; a new form of sharing the urban space between residents and visitors has arisen and a change has occurred in accommodation patterns and in the tourism value chain. Its development is still in the incipient stage but could constitute a key factor in the saturation process or the excessive demand pressure on certain destinations and lead to the discontent of the local population and the rejection of the tourism phenomenon.

This paper addresses the issue using machine learning techniques. Therefore, together with the above question, the second objective of the article is to show how machine learning techniques are especially well-suited for addressing this issue. Thus, a third research question would be:

RQ3 "Is the machine learning methodology useful for addressing these questions?"

More specifically, as an example of potential implementation, it contemplates the creation of an early warning system that will help Spanish destinations manage the potential threat posed by residents and tourists perceiving a problematic situation of overtourism. As far as the authors are aware, this is the first time that this kind of exercise has been attempted and represents an innovation and a step forward with respect to the existing literature on the subject.

The paper is structured as follows. After this introduction, in section two the literature on the topics of destination competitiveness, the sharing economy and overtourism is reviewed. In section three we explain the methodology and the data used for the analysis. The data analysis is performed in section four. In section five we discuss the main results obtained. Finally, section six summarizes the main conclusions and suggests future lines of research.

Literature review

Destination competitiveness: theory, measurement and practice

Competitiveness, a general concept that encompasses price differentials and exchange rate movements, the productivity levels of various components of the tourist industry and qualitative factors that affect the attractiveness or relative desirability of a destination (Dwyer, Forsyth and Rao, 2000), constitutes the basis of tourism success for companies and destinations.

Competitiveness is a relative concept – a destination is competitive in comparison with another destination(s). It is also a complex term, which covers different levels or dimensions (Spence and Hazard, 1988) and for which there are different definitions and measurement methods. Applied to tourist destinations, competitiveness seems to be linked to the capacity of a destination to provide goods and services valued by tourists that are superior to those offered by competing destinations (Dwyer and Kim, 2003). In this sense, the competitiveness models of tourism destinations, as in Crouch and Ritchie (2003) have sought to reflect the attributes that confer competitiveness to destinations from a broad perspective.

But, theoretical models are sometimes difficult to operationalize in practice. Botti and Peypoch (2013) attempt to make the Crouch and Ritchie model operational, showing that competitiveness cannot be understood unilaterally or as a precise value. Likewise, the assessment and decisions made by visitors are completely subjective, depending on the type of destination, or even the type of tourist (Botti and Peypoch, 2013). Even so, the various existing definitions have a common denominator. The improvement of the quality of life of the destination's residents must be the main objective of competitiveness in the tourism sector, or at least be reflected in some way, thus guaranteeing the long-term sustainability of the tourism activity (Crouch and Ritchie, 1999; Dwyer and Kim 2003; Hassan, 2000).

On a theoretical level, competitiveness goes beyond growth or the simple volume of visitors to a destination. If initially the real objective of the policy of promoting the competitiveness of many destinations was to maximize the growth in the number of tourists, now it is acknowledged that this increase in demand only has, theoretically, an instrumental role. Sustaining the activity and improving the quality of life of the residents is the real ultimate objective of competitiveness. In this way, achieving this objective of competitiveness and sustainability can sometimes involve setting limits to growth (Perles, Ramón, Vera and Ivars, 2018) or even the de-growth of the destination itself (Andriotis, 2018; Mansilla and Milano, 2018; Fletcher *et al*, 2019).

However, on a practical level, in the day-to-day management of destinations, many programmes are still clearly oriented towards growth (Dodds and Butler, 2019). In this context, many managers still associate competitiveness with tourism success and in practice seek to achieve rapid growth and a high volume of tourists. Tourism growth and its speed are elements which, as we will see below, are directly linked by the literature to potential overtourism situations. For example, Peeters *et al* (2018:23) include tourism density and tourism intensity in their conceptual model of overtourism as drivers of this phenomenon. Also, the intensification of the traditional policy focused on promoting volume and increases in tourism arrivals are highlighted as causes of overtourism (Peeters *et al*, 2018:27; Dodds and Butler, 2019). All of these factors are traditionally associated with the competitiveness of destinations.

For this reason, in this study it is understood that the level of competitiveness of destinations is an important contributing factor to potential overtourism situations and it is therefore a variable that is introduced in its explanation. In particular, it is expected that the greater the competitiveness, the greater the probability of overtourism, since it would be very difficult to appreciate the opposite phenomenon, a situation of overtourism (understood as an excess of tourism) in an unsuccessful destination.

Regarding the measurement of competitiveness, authors have been adapting to this change in theoretical conception, as well as to the growing complexity that the phenomenon itself has been experiencing over the years. Thus, although the traditional measure par excellence of competitiveness had been market share, there is currently considerable controversy regarding its use (see Perles, Ramón and Sevilla, 2014). Consequently, over the last few years there has been a proliferation of synthetic indicators that attempt to capture the various aspects and determinants that make up the competitiveness of destinations.

Among them, the one that has achieved the greatest international success is the Travel & Tourism Competitiveness Index prepared by the World Economic Forum. Adaptations of this indicator have been made for Spain (Monitur, to measure the competitiveness of Spanish regions and UrbanTur to measure the competitiveness of major cities) by Exceltur.

Specifically, UrbanTur (Exceltur, 2017:32) is a synthetic indicator of the competitiveness of cities that analyses 62 indicators grouped into six pillars. These pillars reflect the capacity of attraction of the offer of leisure products; the capacity of attraction of the offer of business products; competitive conditioning factors of the urban environment and local life; accessibility and mobility; governance and strategic management; and performance, economic and social results.

With regard to specific tourism products, the pillar related to leisure tourism measures aspects associated to diverse products such as shopping and family tourism, events or festivities with potential interest for tourists, cruise tourism, beaches and idiomatic tourism. On the other hand, the pillar referring to business tourism seeks to measure the affluence to fairs and congresses and the endowment of high category hotels and restaurants.

The pillars of the local competitive environment and accessibility and mobility are built on indicators that reflect the attractiveness of tourist spaces and the urban environment, the quality of education and citizen safety, among others. Likewise, air and rail connectivity, pedestrian or bicycle mobility, the efficiency of the taxi service and the sustainability of the city's public transport are also considered.

The governance pillar seeks to reflect the importance of tourism in the local administrative organisation. In addition, promotional efforts in tourism are considered and the application of technologies and intelligence systems to tourism and the raising of awareness of the importance of tourism among citizens. Finally, the pillar relating to the economic results of the destination is based on classic measures such as the average stay, the economic impact of tourism, the profitability of the activity, the social contribution and the market positioning.

Given the availability of these competitiveness measures for the case of Spanish cities, and in accordance with their ability to capture the various aspects or levels of the phenomenon, it is logical that this study has chosen the UrbanTur indicator developed by Exceltur as a variable for reflecting the competitiveness of Spanish urban destinations.

Overtourism, sharing economy and their complexities

Overtourism is usually defined as the "situation in which the impact of tourism, at certain times and in certain locations, exceeds physical, ecological, social, economic, psychological, and/or political capacity thresholds" (Peeters *et al*, 2018:22). UNWTO (2018) defines overtourism as "the impact of tourism on a destination or parts of it, that excessively influence the perceived quality of life of the residents and/or the quality of the experience of the visitors in a negative way".

Overtourism is a problem that has recently erupted with force in some consolidated tourist destinations which has attracted the attention of the media and the scientific community (Huete and Matecon, 2018). However, the problems that tourism generates on the quality of life of the residents have been studied for years and it has been found that they also arise in destinations with small tourist flows (see, for example, Michalkó *et al*., 2013). In Spain, in addition to the well-known case of Barcelona, many other destinations such as Valencia (Bol Esteve and Arnandis-i-Agramunt, 2020), Toledo (Escudero Gómez, 2019), Majorca (Blázquez-Salom *et al*, 2019) and Lanzarote (Carballo *et al*., 2019) have been studied because of the possible existence of excess tourism and its consequences.

The literature conceives it as a complex problem associated with deficiencies in the management of destinations (UNWTO, 2018) that lead to a lack of social sustainability of the activity, which generates a negative attitude among residents (Bellini et. al, 2016; Muler-González, Coromina and Galí, 2018). Furthermore, the literature on overtourism identifies the obsession to continually increase the number of tourists and arrivals as one of the causes of this phenomenon (Séraphin *et al*., 2019; Goodwin, 2017). In this way, the concept gives rise to ramifications connected to a broad range of aspects analyzed by the literature such as competitiveness and the success of destinations (Crouch and Ritchie, 1999; Dwyer and Kim, 2003), their carrying capacity and sustainability (Doxey, 1975; Canestrelli and Costa, 1991; Vera and Ivars, 2003; Perkumienė and Pranskūnienė, 2019), the attitude of residents towards tourism (Doxey, 1975; Getz, 1994; Long and Kayat, 2011; Wang, 2016; Ghasemi, 2019) and the different costs and benefits that the activity generates for the different stakeholders existing in the destination (Nunkoo and Ramkissoon, 2012; Postma and Schmuercker, 2017; Namberger, Jackish, Schmude and Karl, 2019).

Although the difficulties are not exactly new, the emergence of problems deriving from potential situations of overtourism in tourist destinations has occurred simultaneously with the emergence and generalization of what has come to be known as the sharing economy in the field of accommodation (Oklevik *et al*., 2019). This paper refers to this term not only in the original sense in which it was conceived, of sharing or exchanging experiences without the mediation of economic provision, but also in the current sense of having become a disruptive business model whereby platforms are used for the channeling of supply, together with private individuals, companies or the regulated supply of rental accommodation (Ram and Hall, 2018; Volgger and Huang, 2019)

The marketing of apartments and homes through platforms such as Airbnb, driven by the search by tourists for more authentic and interactive experiences (Guttentag, 2015) has undoubtedly had positive effects for the destinations, such as a better optimization of the exploitation of underutilized resources (Martin, 2016). However, in parallel, it has generated negative effects by directing massive flows of tourists to places and areas of residential cities, not designed for tourism functionality. In the absence of adequate management, this has given rise to problems of coexistence and discomfort for residents of these areas who suffer the effects of seasonal tourism peaks (Arbaci and Tapada-Berteli, 2012; Milano *et al*., 2019).

In the same way, the greater profitability of tourist rentals as opposed to residential rentals has led to a significant transfer of properties dedicated to tourism, which has degenerated into price rises and difficulties in accessing housing for the most disadvantaged groups in these areas, inducing gentrification processes and aggravating discontent (Opillard, 2016; Vives-Miró and Rullan, 2017; Alcalde-García *et al*., 2018; Benner 2019). All these reasons lead Peeters *et al* (2018:27) to clearly identify the proliferation of unregulated tourist accommodation as a cause of overtourism and the Airbnb platform as a driver of it. In accordance with this literature, together with the competitiveness of the destinations, this study considers the greater or lesser presence of a supply offered by the collaborative economy in them as another of the fundamental explanatory variables of the potential overtourism situations of the destinations.

The adoption of a supply measure, as opposed to a demand measure, is justified for several reasons. First, because it is easily obtained, given the existence of specialized platforms that allow access to the number of dwellings offered under this system in each destination. Second, in accordance with Say's Law, this supply is expected to be an approximate reflection of the demand in this model supported by the destinations. Although Say's Law is mostly rejected by economists, in the case of mass tourism destinations, the creation of a very wide supply of accommodation was necessary to attract huge flows of demand, which are the cause of the problems related to overtourism. Third, regardless of the degree of occupancy that the dwellings marketed through this method finally reach and their greater or lesser intensity in tourist use, they are former residential properties that have been changed to tourist rentals, so the negative consequences generated by this are already, in any case, real.

Economic dependence and political guidance of destination managers

The role played by the other two explanatory variables included in the model (the economic dependence on tourism in the destinations and the political orientation of their authorities) is, at least for the Spanish case, quite evident. With regard to economic dependence, the greater or lesser tolerance of residents towards the negative externalities generated by tourist activity will depend on the economic development alternatives available in the destination (Ashworth and Page, 2011). In a city with a diversified economy, where the main source of the wealth of its residents is an activity other than tourism (industry, financial services, etc.), there would be a low level of tolerance of the nuisance generated by tourists and it would be more likely to observe an overtourism phenomenon. This is the case of Barcelona, a city where, according to Martín-Martín and others (2018), its residents are well aware of the economic impact of tourism and to a lesser extent it is also beginning to be observed in destinations without as much tourist saturation, such as Porto (Silva *et al*., 2019).

On the contrary, in a city with a high dependence on tourism, where tourism is the main source of wealth and employment for its inhabitants and where there are few or no development alternatives available, this degree of tolerance will necessarily be much greater. It is difficult to observe overtourism in cities or regions that are eminently tourist-oriented, such as in Algarve (Portugal) (Renda *et al*., 2014) or in Benidorm.

Aware of this fact, in their conceptual model, Peeters *et al*. (2018) include tourism's share of Gross Domestic Product (GDP) among the drivers of overtourism. In this article, the economic dependence on tourism is reflected by the average income of households in the region where the destination is located. This is due to the fact that in Spain the municipalities most dependent on tourism are, on the whole, associated to lower per capita income levels of their residents (Plaza, 2018).

With regard to the political guidance from destination managers, in the Spanish case the movements driving the reactions against tourism are mostly social movements that previously fought against the inequalities and problems caused by a neo-liberal economic development (Milano, 2018). Although it is difficult to generalize, these demands and claims of social movements are better received or alienated by local governments, as the authorities themselves sometimes emerge from these movements. Barcelona is perhaps the most representative case of this tendency, but the local governments of other Spanish cities are also concerned about certain overtourism issues, such as Valencia, Eivissa and Palma de Mallorca.

Attitude of residents towards tourism and the measure of overtourism

After identifying competitiveness and the sharing economy as fundamental ingredients of the recipe for overtourism, it is now necessary to examine the attitude of residents towards tourism and how overtourism is measured. This question is related to the objective or subjective approach that can be taken in analysing the phenomenon and that conditions the measurement instruments available to the researcher.

The scarce and preliminary empirical studies on the subject (McKinsey, 2017; Peeters *et al*. 2018) use an objective approach to examining the phenomenon, seeking to collect economic indicators of supply and demand of destinations and other measures of environmental and social pressure that capture the complexity of the phenomenon on which to make comparisons through basic techniques of statistical analysis in search of thresholds. Based on these thresholds, the aim is to detect and inform the diagnoses of the destinations and suggest the decisions and measures to be taken to solve the identified problems.

However, it is also possible to take a subjective approach to the issue, since overtourism is only a reflection of the residents' attitudes towards tourists, which is very much linked to the subjective perceptions of these residents. Peeters *et al* (2018:22) perfectly reflect the subjective nature of overtourism when they point out the key role played in it by the psychological capacity of residents and tourists. Psychological capacity "*refers to the capacity of people (residents and /or other visitors) to emotionally cope with crowding effects*" and is included in the tourism capacity section of their conceptual model of overtourism.

This subjective approach is adopted in this paper. Therefore, as indicated in the methodology and data section, the identification of a potential situation of overtourism in order to create and train the different algorithms object of this study has been based on whether or not the destination belongs to the Network of Southern European Cities against Touristification (SET Network), created in Barcelona on 18 and 19 May 2018 (Peeters *et al*, 2018:30). It includes those cities which, on their own initiative and due to a potential fear of developing problems derived from tourist overcrowding decided to create a forum for dialogue.

It could be argued that the fact that a city belongs to the SET network is a choice of the city which may be related to overtourism but a city seeking to prevent future prospects of overtourism may also wish to enter the network. Thus, being part of the SET Network does not provide a clear definition of the state of overtourism of the city.

However, the network consists not so much of entire cities or the public administrations that represent them, but of neighborhood groups or associations (Moreno and Pardo, 2020). Therefore, the membership of a city or group of the network reflects a situation in which at least some of its citizens perceive that there are problems arising from excessive tourism in the destinations. This is clearly in line with the subjective concept of overtourism advocated in this article.

In any case, this criticism of the indicator could be extended to any other indicator proposed to measure overtourism, including the objective indicators that have been proposed to date by the existing literature on overtourism. This is because the existing objective indicators reflect overcrowding situations, which may or may not lead to overtourism situations. In this sense, and in the absence of an obvious and generally accepted indicator of the existence of a situation of overtourism, the belonging of the city to the SET Network could be an acceptable measure of overtourism.

Having examined the literature on the phenomenon and having justified the relevant variables and measurements adopted, Figure 1 shows the proposed theoretical model of destination overtourism. On the left-hand side of the figure are the causes of overtourism that have been discussed in the literature review. An excessive demand pressure on the destination, a reflection of a misunderstood excess competitiveness - i.e. based solely on volume and growth targets- and a rapid growth in the collaborative accommodation supply are necessary causes, but not sufficient to generate the emergence of a situation of overtourism on destinations.

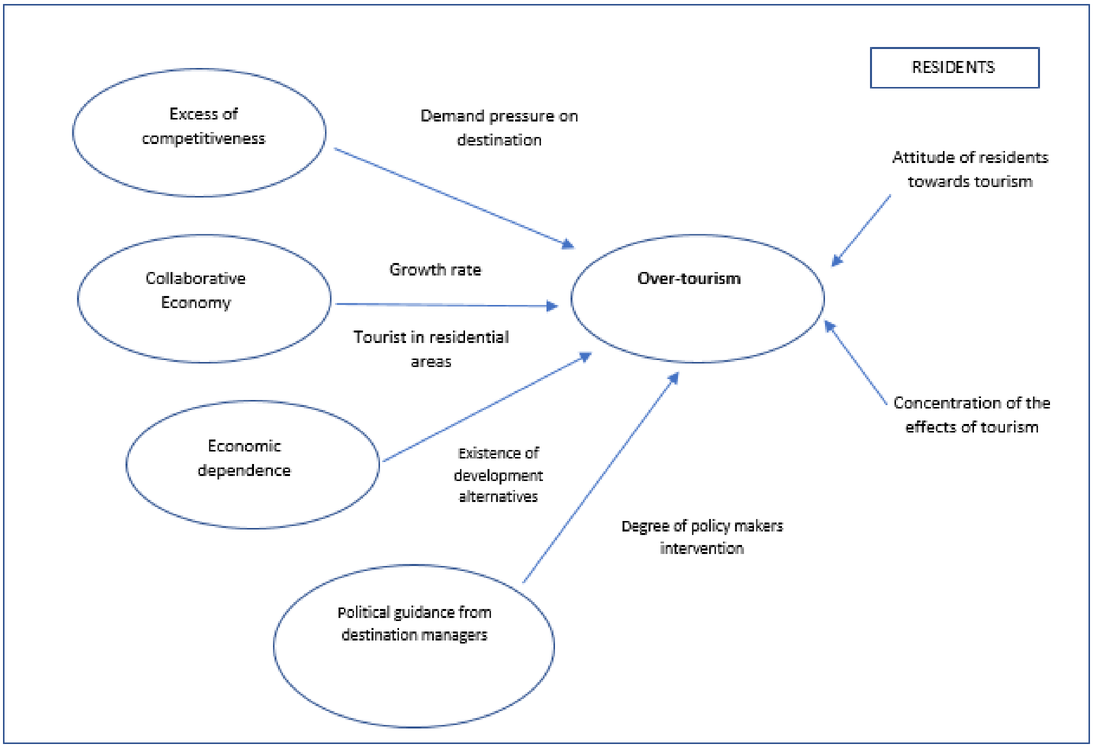


Figure 1. *Proposed theoretical model on destinations overtourism*

Whether these causes translate into a real situation of overtourism will ultimately depend on the degree of dependency that the destination has on the tourism activity (the greater the diversification and lesser the dependency, the greater the probability of a situation of overtourism) and on the greater or lesser receptiveness of the authorities of the destination to the demands of the most sensitive social groups on the subject.

Finally, the strength and intensity of these anti-tourism groups will depend, in turn, on the more or less tolerant attitude of the residents towards the tourism activity (located on the left-hand side of the figure), which in turn will be influenced by the degree of concentration of the negative externalities of tourism in the territory. The greater the concentration, the greater the likelihood of a negative reaction from the residents.

The last section of the literature review defends the suitability of Spain as a case study for the phenomenon, emphasizing the heterogeneity of situations existing between the different regions and tourist cities that make up the Spanish reality.

Spain as a competitive tourism destination: a case study

According to the latest reports published by the World Economic Forum, Spain has held privileged positions in the competitiveness ranking of world tourist destinations for more than a decade. Moreover, since 2015 it has been the global leader, thanks mainly to its outstanding position in aspects related to infrastructure and natural and heritage resources (see Table 1).

Table 1. *Spain's position in the 2015-2019 WEF-TTCI competitiveness sub-indices*

Year	2015	2017	2019
Number of countries considered	141	136	140
Overall Rank	1	1	1
Enabling Environment	35	30	33
Travel & Tourism Policy and Enabling Conditions	8	7	10
Infrastructure	2	4	4
Natural and Cultural Resources	4	3	3

Source: World Economic Forum (WEF) 2015, 2017, 2019.

Spain’s world leadership position does not imply that the country is a homogeneous tourism reality. On the contrary, within the country there are important differences at a regional (see Table 2) or local (see Table 3) level, with each territory having competitive advantages and disadvantages depending on the evolution of the situation itself, the greater or lesser difficulties for accessibility and the decisions adopted in terms of tourism policy by those responsible.

This different competitive situation conditions the response of each territory and each city to the disruptive phenomena recently experienced by tourist destinations. This is particularly the case of the sharing economy, which is seen differently and gives rise to potential overtourism situations, depending on the greater or lesser capacity to manage the impacts and effects it generates.

In the specific case of the large cities representing urban tourism, the unit of analysis in this article, Barcelona, Madrid and Valencia (all of them belong to the SET Network) occupy the top positions in the urban tourist destinations competitiveness ranking in Spain. This competitiveness, understood as the capacity to attract tourists, has led to a sharp increase in the supply of Airbnb accommodation and certain effects such as gentrification, the discomfort of neighbours or increased land prices (Adamiak *et al*, 2019; Santos and Sequera, 2018), which could be related to situations of overtourism.

Table 2. *Competitiveness of the Spanish regions*

Destination	Rank 2014	Index 2014	Rank 2018	Index 2018	Rank variation
Andalucía	7	101.2	6	107.2	1
Aragón	13	94.0	17	91.1	-4
Asturias	14	93.1	11	95.0	3
Balears, Illes	5	103.2	7	106.3	-2
C. Valenciana	6	101.5	4	108.6	2
Canarias	4	103.8	2	111.2	2
Cantabria	12	94.7	14	92.8	-2
C. La Mancha	15	91.8	13	93.6	2
Castilla y León	11	99.8	8	97.0	3
Cataluña	1	116.0	5	108.3	-4
Extremadura	17	90.3	16	91.2	1
Galicia	10	100.0	9	95.9	1
Madrid	2	112.4	3	109.9	-1
Murcia	16	91.7	12	94.9	4
Navarra	8	100.1	10	95.2	-2
País Vasco	3	110.4	1	112.6	2
Rioja, La	9	100.1	15	91.4	-6

Source: Exceltur (2019)

These are cities that combine various products, and although cultural and business trips predominate, the accommodation offered on platforms such as Airbnb is segmented into entire place (generally intended for long stays and several tourists) and private room (more related to a business or student tourist, or very short stays of a cultural nature) (Moreno-Izquierdo *et al*, 2019).

On the other hand, coastal cities with a greater component of the sun and beach model (the Spanish product par excellence) occupy much more modest positions in this ranking of competitiveness. This would be the case of Palma de Mallorca, Las Palmas de Gran Canaria, Málaga, Alicante or Murcia. Some of these cities belong to the SET Network (Palma de Mallorca or Las Palmas de Gran Canaria, Malaga) and others do not (Alicante and Murcia). In this type of destination, the offer of platforms such as Airbnb is more specialized in entire apartments, mainly due to the high number of second homes on the Spanish coast (Moreno-Izquierdo *et al*, 2019). In addition, in these cities, there is a lower incidence of non-holiday tourism, so the pressure and effects of overtourism could be limited to specific times of the year. However, such effects could be sufficient to provoke an overtourism effect in the destinations, as Stanchev (2018) explains in the case of Mallorca, and which could be replicated in the rest of the coastal destinations.

Finally, the medium-sized cities of the interior of the peninsula with a tradition in cultural tourism, together with certain cities of the Cantabrian coast and the Atlantic slope, also offering a cultural or gastronomic product, are placed in the lower positions of the competitiveness ranking. So, although Airbnb and similar platforms offer listings in the destinations, these hardly have any impact on the pressure of the cities or the feeling of overtourism by the citizens. A detailed examination shows that, on the whole, these cities do not belong to the network.

Table 3. *Competitiveness of the main Spanish urban destinations*

Destination	Rank 2012	Rank 2016	Index 2012 (average=100)	Index 2016 (average=100)
Barcelona	1	1	141.40	145.80
Madrid	2	2	139.10	138.00
Valencia	3	3	111.00	108.20
San Sebastián	5	4	103.10	104.50
Málaga	6	5	98.30	103.00
Sevilla	4	6	104.30	102.20
Palma de Mallorca		7		101.50
Bilbao	7	8	97.90	99.20
Santiago de Compostela	8	9	97.30	96.40
Gijón	12	10	94.20	96.30
Las Palmas de Gran Canaria		11		96.20
Valladolid		12		95.60
Zaragoza	9	13	96.60	95.50
Córdoba	13	14	93.90	95.30
Santander	16	15	91.10	95.20
Granada	10	16	95.90	91.60
Salamanca	11	17	95.70	91.50
Alicante	15	18	91.90	91.30
Murcia		19		89.50
La Coruña	14	20	92.40	88.80
Burgos	18	21	88.00	87.70
León	20	22	85.70	85.80
Toledo	17		90.80	
Oviedo	19		86.40	

Source: Exceltur (2013, 2017)

Methodology and data

According to the analysis of the literature, the basic model could be formally described as OV=f(COM, PRES, DEP) where OV is an indicator of a situation of overtourism (real or perceived) in the destination, COM represents its competitiveness, PRES is a variable reflecting the degree of pressure exerted by the tourism industry on the destination and DEP is the dependence of the destination on the tourism industry. In order to test whether the sharing economy or the traditional offer have different effects on overtourism, the pressure variable is divided into the two components: SH is the variable that reflects the pressure from the sharing economy, PRESTRAD reflects the pressure from the traditional offer.

Regardless of the functional form of the model and seeing the phenomenon of overtourism as an excess of tourism competitiveness (destinations victim of one's owns success), it is expected that the competitiveness (COM) and pressure variables (PRESTAD and SH) should be positively (directly) associated with the overtourism indicator (OV). Conversely, the tourism dependence variables should be negatively (inversely) associated with the overtourism indicator.

As stated in the introduction section, this paper addresses the hypotheses testing using machine learning techniques. After exploring the adequacy of the data for the planned techniques using exploratory data analysis (EDA), several techniques linked to binary classification are used. First, a logit model is estimated and deeply explored as a basis for the analysis. Second, in order to check the robustness of the results obtained, several alternative techniques are used such as Naïve Bayes (NB) classification; Support Vector Machines (SVM) classification and regression tree (RT).

The use and comparison of several techniques is due to the exploratory nature of the article and the aim to show how machine learning is especially well-suited to addressing this topic. Eventually, in other kinds of paper (more advanced exercises that attempt to establish definitive results), the recommended methodology would be to compare the performance of each technique for the data at hand using their predictive capacity based on the metrics shown in Table 8, and select the best one to conduct the predictive exercise for new destinations. Likewise, it would be possible to use other available machine learning techniques in order to perform the exercise.

According to Chollet and Allaire (2018), machine learning is a subfield of artificial intelligence in which computers automatically learn the rules governing the data analyzed, rather than programmers crafting data-processing by hand. With machine learning "*humans input data as well as the answers expected from the data, and outcome the rules. These rules can be applied to new data to produce original answers*" (Chollet and Allaire, 2018:5). In this sense, a machine learning system is trained rather than explicitly programmed.

Machine learning models try to find appropriate representations for their input data searching through a predefined set of operations, called a hypothesis space. The objective of these data transformations is to make the task at hand, such as a classification task, more amenable. The techniques used in this article are included within the so-called classification techniques. A brief description of the specific techniques used in this article can be seen in Chollet and Allaire (2018).

While Bayesian classifiers and regression trees are considered as basic techniques of data classification, Support Vector Machines are included in the advanced techniques category. Han, Kamber and Pei, (2012) is used as a reference to summarize the applied techniques in this article. A common characteristic of the techniques used in this study is that they involve a two-step process: the first is the construction of the classification model (learning step or training phase) using a training set of the data; the second process is that of prediction based on the constructed model using a test set of the data (classification step). The techniques used fall within the category of supervised machine learning given that for the construction and validation of the model, the class label of each training tuple (outcome category) is known.

Once the model is trained, the predictive accuracy of the classifier is estimated using the test set. The final goal of this phase is to avoid a model that tends to overfit the data. Many metrics to evaluate the classifier performance exist, but the most elementary may be the accuracy of a classifier on a given test set, defined as the percentage of test tuples that are correctly classified by the model.

Beginning with the basic techniques, Bayesian classifiers are statistical classifiers based on Baye's theorem. The most relevant feature is that Naïve Bayesian classifiers assume that the effect of an attribute value on a given class is independent of the values of the other attributes.

$$P(X|C_i) = \prod_{k=1}^{n} P(C_i) = P(C_i)xP(C_i)x \dots xP(x_n|C_i)$$

Given a tuple "X" (in our context, a vector with the destination features), the classifier will predict that "X" belongs to the class having the highest posterior probability, conditioned on "X".

$$P(C_i)P(C_i) > P\big(C_j\big)\, for\ 1 \le j \le m, j \ne i$$

In theory, Bayesian classifiers have the minimum error rate in comparison to other classifiers. However, in practice this is not always the case in empirical work, due to inaccuracies in the assumptions made for their use.

On the other hand, classification and regression tree (CART) dates back to the late 1970s. Han, Kamber and Pei, (2012:330) defines decision trees as "a flow-chart-like structure, where each internal node (noon-leaf node) denotes a test on an attribute, each brand represents an outcome of the test, and each leaf node (or terminal node) holds a class label". When a decision tree is built, some branches reflect anomalies in the training data (outliers), so the tree is usually pruned in order to avoid overfitting issues.

There are several algorithms to build decision trees. Most of them follow a top-down approach. The training set is recursively partitioned into smaller subsets as the tree is being built. The recursive portioning is based on criteria and measures such as information gain or the Gini index. The recursive partitioning stops only when 1) all the tuples in a partition D (represented at a node N) belong to the same class; or 2) there are no remaining attributes on which tuples may be further partitioned; or 3) there are no tuples for a given branch, that is, a partition D_j is empty.

Finally, the support vector machine (SVM) is an algorithm that uses nonlinear mapping to transform the original training data into a higher dimension. Within this new dimension, the linear optimal separating hyperplane is sought: a decision boundary that separates the tuples of one class from another. With an appropriate nonlinear mapping to a sufficiently high dimension, two types of data can always be separated by a hyperplane. An attractive feature of this method is that it is much less prone to overfitting than other methods. According to Chollet and Allaire (2018:16) "at the time they were developed SVMs exhibited state-or-the-art performance on simple classification problems and were one of the few machine learning methods backed by extensive theory and amenable to serious mathematical analysis, making them well understood and easily interpretable".

The estimations in this paper are carried out using the CARET-Classification and Regression Training (Khun, 2020) package for the R 4.0.1 programming language (R Core Team, 2020).

Our data set comprises a pool of 21 cities analysed by Exceltur in the Monitor competitiveness of Spanish urban destinations. As explained in the model and in Table 4, an indicator variable is used as a dependent variable that reaches 1 if the corresponding city belongs to the Red de Ciudades contra la masificación turística (Network of Cities in response to the Massification of Tourism) or not. The cities belonging to this network are Barcelona, Madrid, Valencia, San Sebastián, Málaga, Sevilla, Palma de Mallorca and Las Palmas de Gran Canaria. The rest of the cities that do not belong to the network are

Bilbao, Santiago de Compostela, Gijon, Valladolid, Zaragoza, Cordoba, Santander, Granada, Salamanca, Alicante, Murcia, La Coruña and Burgos.

Table 4. *Data and sources*

Variable	Description	Value	Source
OV	Overtourism, 1 if the city is included in the network against tourism massification; 0 if not.	2017	Peeters et. al (2018)
COM	Competitiveness Index	2016	Exceltur (2016)
PRESTRAD	Ratio between hotel overnight stays and the city's population	2016	Spanish National Statistics Institute (INE 2017a).
SH	Ratio between Airbnb accommodation supply (2018) and the city's population (2016)	2018	AirDNA, Spanish National Statistics Institute (INE 2017b)
DEP	Average income of households in the region where the destination is located	2014	INE (2016)

It could be argued that the small sample size would be an important limitation to test the theoretical model. However, the universe of the study comprises the urban destinations in Spain. In this sense, the 21 cases of the sample considered, being regional capitals and major cities, are highly relevant in the context of Spanish urban destinations. In fact, it could even be considered that the sample could constitute the entire study universe, in which case, following the line of reasoning of Ziliak and McCloskey (2008) statistical significance would not be a relevant issue in this context. Without assuming such an extreme position, the authors consider for the case at hand and again bearing in mind the exploratory nature of this article that “data relevance beats data size”, thus the analytical and inference exercise carried out would be correct from a methodological and statistical viewpoint.

It could also be argued that the machine learning technique is a useless complication for treating a very simple case comprising 21 observations that could be analysed using basic descriptive techniques. On the one hand, as stated in the introduction section, one of the objectives of the paper is to show how machine learning techniques are especially well-suited to addressing overtourism topics. And, as usual in these kinds of articles that propose and apply methodological innovations to a new field, in order to assess the goodness of the methodology it is necessary to use a dataset that is easily manageable and interpretable, which is the case of the sample selected in this article. On the other hand, the fact that machine learning techniques are usually designed to perform their task in a big data context does not prevent their application in a small data situation. However, as recommended by Zhang and Ling (2018), in this case it is necessary to acknowledge the mediation relationship existing between the data size, model complexity and predictive performance of the models, and to prevent some kind of strategy to improve this performance. With respect to the tourism literature, Perles, Ramón, Moreno and Sevilla (2016) is an example of the application of machine learning techniques to a context of a small dataset.

In this article, so as to deal with such a small and imbalanced sample, cross-validation techniques and upsampling methods have been applied to improve the predictive properties of the estimated models. Specifically, in order to assess the predictive capacity of the model, the dataset has been split into two groups, 70% of the data are devoted to the estimation of the model (training dataset) and the remaining 30% to its validation (testing dataset).

Although the imbalance of the sample is not very extreme (8 positive versus 13 negative cases), before training the models and following the recommendations of Khun and Johnson (2013:439), an upsampling has been performed by increasing the number of non-overtourism cities using the upSample function of the caret package. In this way the overtourism and non-overtourism cities in the training dataset have been balanced. Also, to train the models, 10 fold cross-validation repeated 10 times has been used as a control parameter for the train caret estimation function. Finally, definitive predictive performance measures have been calculated using the predictions of the trained model on the testing dataset.

Results

Exploratory data analysis

Table 5 shows the descriptive statistics for the variables considered in the model and the size of the population in the cities considered for the analysis and Table 6 shows the correlation matrix that suggests no severe problems of collinearity are implied from the observation of linear correlations. It can be seen that cities differ in all of the variables with cities with overtourism issues reaching higher values in all variables.

Table 5. *Descriptive statistics*

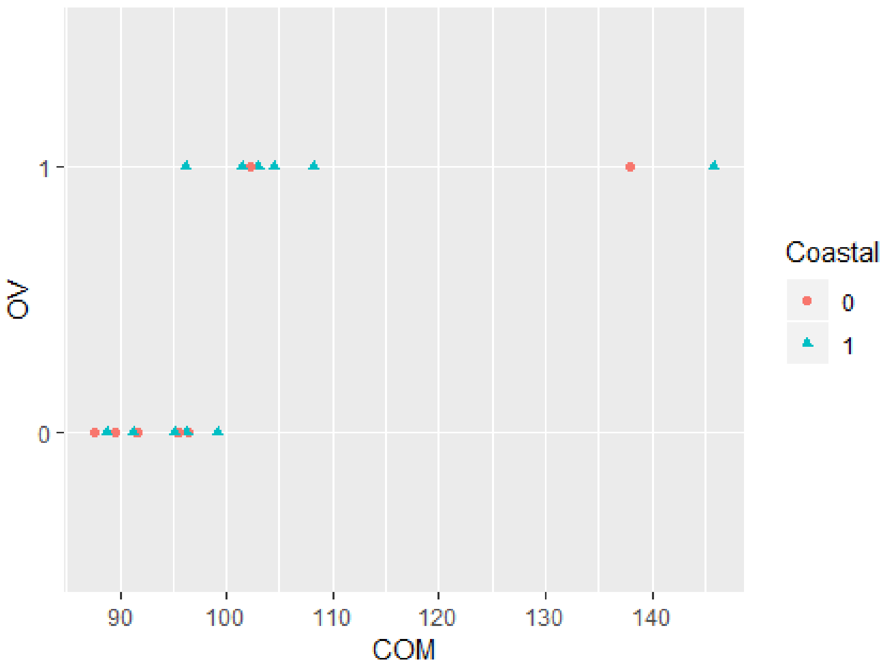
Variable	**Whole sample average (standard deviation) N=21**	**OV=1 average (standard deviation) N=8**	**OV=0 average (standard deviation) N=13**
COM	100.6 (14.77)	112.42 (18.61)	93.38 (3.50)
PRESTRAD	6.59 (4.89)	8.20 (5.93)	5.60 (4.09)
SH	5.23 (3.72)	7.56 (2.44)	3.79 (3.70)
DEP	29798 (3830)	31144 (5130)	28968 (2674)
Population	549555 (684489)	974009 (984152)	288352 (146461)

Table 6. *Pearson correlation table*

	OV	PRESTRAD	SH	COM	DEP
OV	1.00				
PRESTRAD	0.26	1.00			
SH	0.50	0.46	1.00		
COM	0.64	0.23	0.40	1.00	
DEP	0.28	0.19	-0.06	0.61	1.00

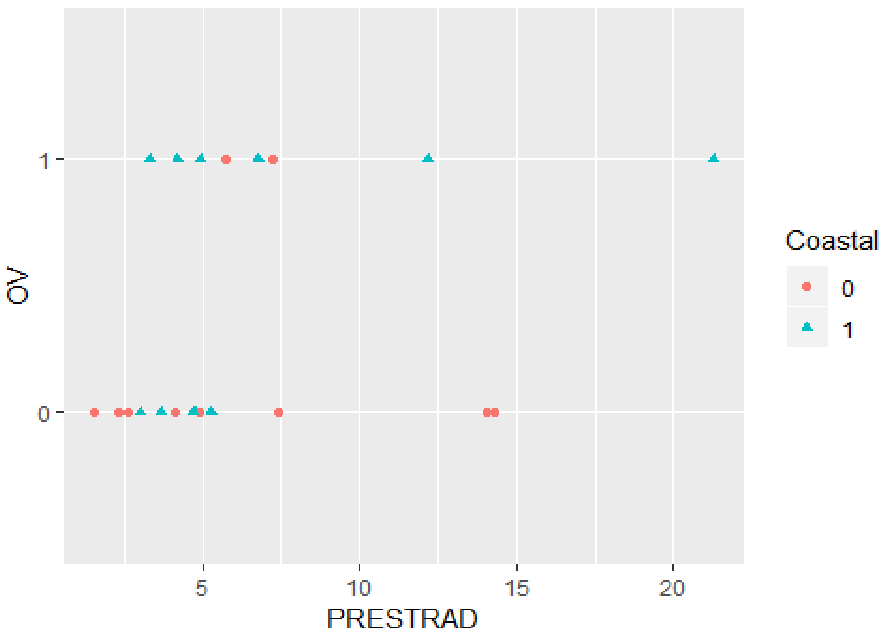
Figure 2 shows the relationship existing between the overtourism indicator and the competitiveness index. It may be observed that a perfect separation practically exists between the two groups, which could be seen (and referred to later) as a difficulty for model estimations. Figure 3 and 4 show the relationship existing between the overtourism indicator and the pressure variables. A greater mix can be observed between the destination groups, indicating that the relevant variable in this context is the competitiveness index. However, for the variable SH, except in the case of certain outliers, there is also a high degree of separation between destinations.

Finally, Figure 5 shows the relationship existing between the overtourism indicator and the dependence (income population) variable. As in the case for pressure variables, a greater mix may be observed between the destination groups.



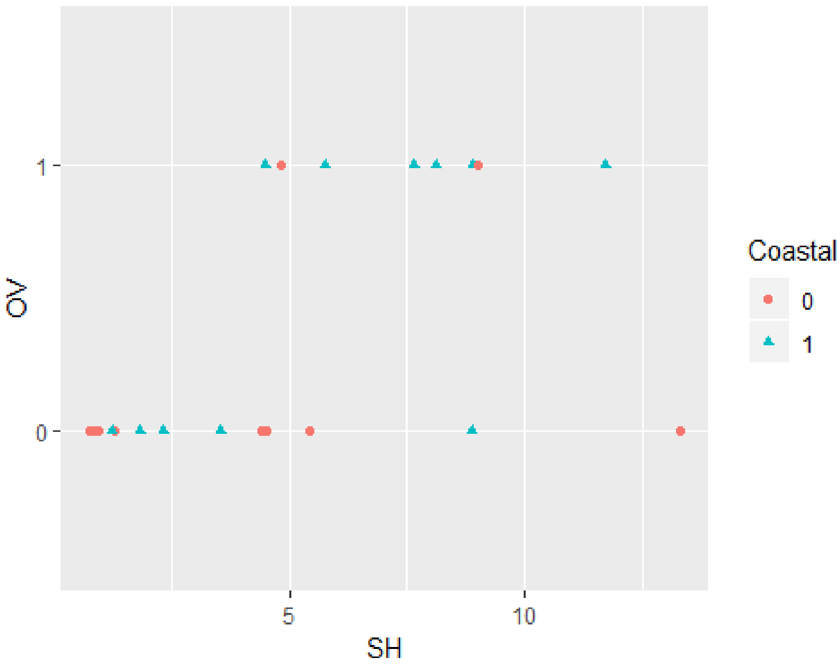
Note: COM Competitiveness index; Non overtourism destinations (OV=0), overtourism destinations (OV=1)

Figure 2. *Relationship existing between competitiveness in destinations and overtourism*



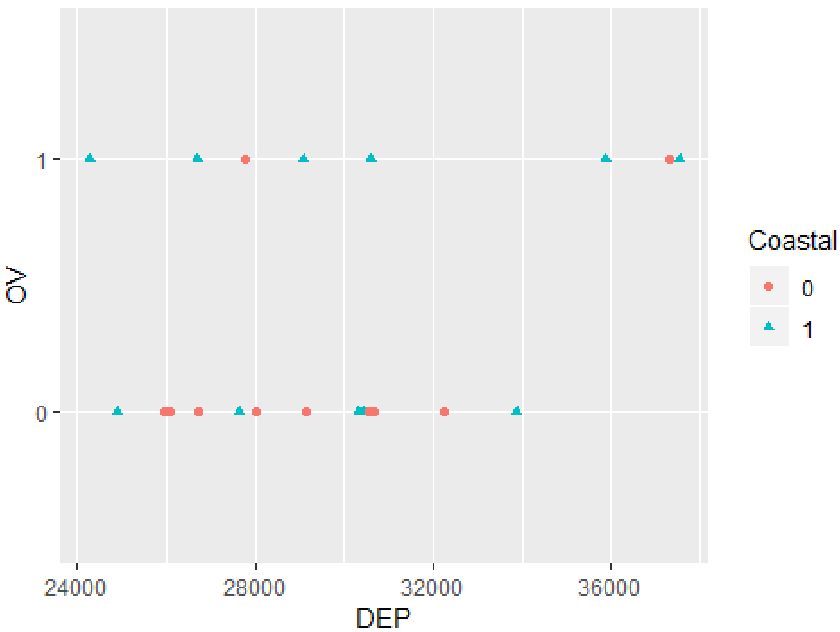
Note: PRESTRAD Traditional accommodation pressure in destinations; Non overtourism destinations (OV=0), overtourism destinations (OV=1)

Figure 3. *Relationship existing between traditional accommodation supply and overtourism in destinations.*



Note: SH Sharing economy accommodation pressure in destinations; Non overtourism destinations (OV=0), overtourism destinations (OV=1)

Figure 4. *Relationship existing between sharing economy accommodation supply and overtourism in destinations.*



Note: DEP economic dependence of tourism in destinations; Non overtourism destinations (OV=0), overtourism destinations (OV=1)

Figure 5. *Relationship between tourism economic dependence and overtourism in destinations*

A classification based on the Logistic regression model

In the original training dataset, five of the 15 cases (33.00 percent) are positive cases (overtourism cities), thus, as a baseline, any method that achieves a level of accuracy of more than 66 percent would be useful. As stated above, the first estimation of the model considered is through a logistic regression model (logit), with the form:

$$\text{Logit (OV)} = \beta_0+\beta_1\text{COM}+\beta_2\text{PRESTRAD}+\beta_3\text{SH}+\beta_4\text{DEP}$$

Table 7 reflects the result of a Bayesian estimation through the caret package using the bglm function.

Table 7. *Bayesian estimation*

Coefficients	Estimate	Std. Error	Z value	P-value
(Intercept)	-72.289602	33.047226	-2.19	0.029 *
COM	0.789377	0.344702	2.29	0.022 *
DEP	-0.000164	0.000274	-0.60	0.550
PRESTRAD	-0.045474	0.158305	-0.29	0.774
SH	0.249034	0.228803	1.09	0.276

Levels of significance: *** p<0.001, **p< 0.01, *p< 0.05
(Dispersion parameter for binomial family taken to be 1)
Null deviance: 27.7259 df= 19
Residual deviance: 5.5117 df= 15
AIC: 15.51 Number of Fisher Scoring iterations: 53

The result obtained partially coincides with the previous expectations; the competitiveness has a positive sign. Pressure coming from the sharing economy presents a positive sign, in accordance with expectations. However, pressure coming from traditional accommodation presents a negative sign suggesting that this segment of the demand does not appear to be responsible for overtourism issues. Conversely, the dependence variable is negatively associated with the overtourism indicator.

With regard to statistical significance, although with only 21 observations it is not a concern in this case, it can be appreciated that, apart from the constant, only the competitiveness index reaches statistical significance. Therefore, everything points to the fact that in the case of Spanish urban destinations, the determining factor of a situation of overtourism, real or perceived, is the competitiveness of these destinations. This is the hypothesis of this study.

Once the model is estimated, it is necessary to assess its accuracy. This is done by exploring the measures in Tables 8 and 9. According to Chollet and Allaire (2018), the most relevant for the small data at hand are accuracy, sensitivity and specificity. Accuracy is a summary measure of the correctly classified cases (positive and negative). Sensitivity (also called the true positive rate, the recall or probability of detection) measures the proportion of actual positive cases (overtourism cities) that are correctly identified as such. Finally, specificity (also called the true negative rate) measures the proportion of negative cases (non-overtourism cities) that are correctly identified as such.

Setting the threshold of classification at 0.5, the predictive capacity values of the model that can be observed in Table 7 are obtained. The model misclassifies only one of the six cases of the testing data set, reaching an accuracy level of 0.83. This case is a negative one (i.e. one of the six destinations with non-overtourism issues) with a value of the sensitivity (true positive recognition) of 1 and specificity (true negative recognition) of 0.66.

Table 8. *Predictive capacity of the logit model*

Measures	Cross-validation measures	In testing dataset
Accuracy	0.85	0.83
Kappa	0.71	0.66
Sensitivity	0.81	1.00
Specificity	0.90	0.66

Note: Cross-validation measures presented for the best tuning parameters.

Alternative classification methods
In order to check the robustness of the results obtained, apart from the logistic regression other machine learning techniques have been used to estimate the model.

Table 9 shows this in terms of accuracy. Using the cross-validation measures, BN performs the best and outperforms the logistic regression and RT performs the worst. Using the measures obtained through the testing dataset, two of the three alternative techniques (SVM and BN) equal the result of the logistic regression model and only CART performs worse than the logit model and the baseline naive prediction. This shows the appropriateness of any of them to address the problem analysed based on the available data.

Table 9. *Predictive capacity of SVM, BN and CART methods*

Measure	Classification and Regression tree (CART)		Bayes Naïve (BN)		Support Vector Machine (SVM)	
Measures	**CV**	**Testing**	**CV**	**Testing**	**CV**	**Testing**
Accuracy	0.50	0.50	0.91	0.83	0.75	0.83
Kappa	0.00	0.00	0.82	0.66	0.51	0.66
Sensitivity	1.00	0.00	0.82	0.66	0.83	1.00
Specificity	0.00	1.00	1.00	1.00	0.68	0.66

Note: Cross-validation measures presented for the best tuning parameters.

Discussion

The main result shown in the previous section is the existence of a direct and statistically significant relationship between the level of competitiveness of destinations and their probability of suffering an overtourism situation. Apparently, this result challenges the general conception seen in the literature review (Crouch and Ritchie, 1999; Dwyer, Forsyth and Rao, 2000; Dwyer and Kim 2003; Hassan, 2000) that increased competitiveness always has a beneficial effect on destinations, indicating that overtourism may reflect an excess of competitiveness in these destinations (Peeters *et al.*, 2018).

The explanation for this result can come from three potential sources. A first explanation would relate to the practical difficulties in measuring competitiveness, finding a good proxy for it (Perles, Ramón and Sevilla, 2014). An erroneous choice of the variable that reflects competitiveness, and which gives excessive importance to the volume of demand or market share achieved by the destination, could explain a counter-intuitive result such as the one obtained: that increases in competitiveness can lead to overtourism. However, as has been seen in the literature review, this does not seem to be the case here. The UrbanTur competitiveness index (Exceltur, 2017) is similar to the most internationally accepted tourism competitiveness indicators. Therefore, the measurement error in the competitiveness variable would not, in principle, explain the result.

The second explanation would come from actual destination management practice that would not be captured by competitiveness indicators. If, as has been stated in the literature review, on a practical level, many actions to promote competitiveness are growth-oriented, the success of these measures could generate diseconomies of concentration that exceed the advantages obtained from tourism activity, leading to discontent among residents and increasing the risk of overtourism. This result is in line with the suggestions of Séraphin *et al* (2019), Goodwin (2017) and Peeters *et al* (2018) and the corrective efforts needed in this context according to Perles, Ramón, Vera and Ivars (2018), Andriotis (2018) and Fletcher *et al* (2019).

This line of reasoning would also be validated by the coefficients obtained in the model on the variables representing the demand pressure (both traditional and from the collaborative economy) existing in destinations. Although neither reaches statistical significance, both point to the existence of a direct relationship between pressure and overtourism. This issue could be related to the complex and ambiguous interactions that, especially with regard to the collaborative economy (Oklevik *et al*., 2019), occur among the residents of the destination (Alcalde-García *et al*., 2018; Arbaci and Tapada-Berteli, 2012; Guttentag, 2015; Opillard, 2016).

The third explanation would lead to a discussion of the role of the mechanisms for distributing the benefits generated by tourism in the destination, their influence on the quality of life of the residents and therefore the attitude of the residents towards tourism (Nunkoo and Ramkissoon, 2012; Postma and Schmuercker, 2017; Namberger, Jackish, Schmude and Karl, 2019). In the case of tourism dependence, the type of accommodation and its impact on the welfare of residents is relevant (Favre, 2017; Perles, Ramón; Moreno, Such, 2020). The quality of life of the residents can be partly associated with the variable that reflects the economic dependence of the destination.

Its outcome is ambiguous. Its negative sign indicates that higher levels of per capita income in the destinations are associated with higher levels of welfare generated by tourism and a lower probability of overtourism. However, as has also been pointed out in the literature review, higher levels of income would reflect, at least in the Spanish case, less economic dependence on tourism (Plaza, 2018), so a positive coefficient would also have been plausible in the estimate. The lack of statistical significance of the coefficient leaves both possibilities open.

Conclusion

Contribution

This article seeks to shed light on a subject that is becoming a growing concern for academics and managers of tourist destinations: overtourism. The article does so in an innovative way, applying modern machine learning techniques to a case study; Spanish urban destinations. Specifically, and in accordance with the previous literature, one hypothesis is tested, related to the perception of the overtourism phenomenon as a reflection of an excess of competitiveness.

After devising a potential model and testing it with the available data for the Spanish case, the result obtained favours the veracity of the hypothesis. In this sense, all the analyses carried out indicate that, at least for Spanish cities, the competitiveness index is the relevant variable in this context. This variable reaches statistical significance in the logistic regression model conducted. Furthermore, competitiveness is the only variable that appears in the estimated classification and regression tree.

On the other hand, the proxy reflecting the sharing economy accommodation supply appears in the logistic model with the correct sign, but this variable does not reach statistical significance. The estimations made with alternative methods produce good results in terms of accuracy, which shows that this methodology based on machine learning techniques is appropriate for addressing the problem at hand.

Policy implications

In view of this result, it could be argued that the policy implications of this would be disastrous for tourism development as it follows that destinations should not become “competitive” or “successful” beyond a certain threshold level as this may lead to overtourism defined as a decline in the quality of life of residents and/or the quality of experience for visitors. This could not be further from the truth.

The management implications for the destinations in this article are related to the possibility of a "death by success" that a rapid growth of the demand reflecting a high competitiveness can provoke in the destinations. What is advocated in this article is the defence of a broad vision of competitiveness (not based solely on mere economic growth objectives) and its integration with the long-term sustainability (environmental, but also economic and social) of the destination that places the improvement of the quality of life of the destination's residents at the centre. This broad vision of competitiveness must be considered within the framework of concrete actions to promote it, following different methodologies such as the one described in the study by Botti and Peypock (2013)

In this sense, managers of successful destinations must take extreme precaution, monitoring the impact that tourism is generating in the destination and the perception of local residents in order to prevent any negative attitudes that they may have from giving rise to a situation of overtourism with negative consequences for their own competitiveness.

At this point, one of the most sensitive aspects is the distribution of the benefits generated by the tourism activity itself, the economic impacts, in terms of income and employment. The available evidence suggests that obtaining benefits from the tourism activity is fundamental for having a favourable perception of tourism, in destinations in both advanced and emerging countries. Therefore, whether there are more or fewer people in the destination who benefit directly or indirectly from tourism is crucial to achieve a greater or lesser acceptance of the phenomenon in the destination.

Apart from the above, measures such as the diversification of tourism activities, the management of demand flows and the control and management of the commercial activity of collaborative economy platforms are potential measures to be adopted to alleviate situations of overtourism. A further measure is the application of technology to destination management within the framework of what in Spain has come to be called Intelligent Tourist Destinations or Smart Sustainability (Perles and Ivars, 2019) proposals based on a true governance of the destination and is aimed at taking advantage of the opportunities offered by information and communication technologies (ICTs) for a more efficient and sustainable management. The implementation of these measures may possibly have mitigating effects on competitiveness but may render the destination less prone to overtourism.

Apart from warning about the effects that a misunderstood competitiveness concept (based only on growth objectives) could have on destinations, the possibility of creating an early warning system as a practical tool for destination managers based on machine learning techniques is the other main contribution of the article. Machine learning techniques would probably identify potential thresholds in the relevant variables and determinants of overtourism in future case studies that have difficulty in identifying with the statistical descriptive techniques used by the literature to date.

Limitations and future research directions

The article has some limitations. The small sample size is one of the most relevant. Likewise, the analysis carried out leaves out some potentially relevant variables, such as the political orientation of the managers of the destinations. In the case of Spanish cities, some of them (Barcelona, Palma de Mallorca or Madrid) have already shown signs of overtourism within the framework of a relevant political instrumentation (Huete and Mantecón, 2018). Finally, the proxy used for the overtourism situation of destinations is also a limitation for the study carried out. In this sense, the creation of indicators and measures reflecting these complex issues is always a challenge for researchers.

The above limitations are closely associated with the future lines of research suggested by the article. These would be, first, to broaden the universe of the destinations analysed, to include not only urban destinations, but also sun and beach destinations. It is common knowledge that these destinations constitute the main tourist product of Spain. Second, the consideration of additional aspects and variables that have not been taken into account in this initial study, such as the role that each determinant of competitiveness (basic or advanced) plays in potential overtourism situations, the effect that its reputation in social networks has on the attractiveness of each destination, as well as the possible externalities that these may provoke, the role played by seasonality or the rate of growth in demand. And, third, the mitigating role that intelligence or technology applied to the management of each destination can play in the prevention or correction of potential overtourism situations.

The enlargement of the sample, and if necessary the obtaining of longitudinal data, would allow improvements to be made in the methodology and the application of techniques not explored in this article, such as neural network or panel data analysis, which would perhaps enable the results and conclusions to be refined with respect to this preliminary work.

In any case, the exercise carried out represents an attempt (probably the first) to address these questions using machine learning techniques and represents a step forward with respect to the existing literature on the subject.

References

Adamiak, C., Szyda, B., Dubownik, A., and García-Álvarez, D. (2019). Airbnb offer in Spain—spatial analysis of the pattern and determinants of its distribution. *ISPRS international journal of geo-information* 8(3) 155, 1-26

Alcalde-García, J., Guitart-Casalderrey, N., Pitarch-March, A. and Vallvé-Fernández, O. (2018). De la turismofobia a la convivencia turística: el caso de Barcelona. Análisis comparativo con Ámsterdam y Berlín. *Revista ARA* 8 (2), 25-34

Andriotis, K. (2018). *Degrowth in Tourism. Conceptual, Theoretical and Philosophical Issues*. London: CABI Publishing.

Arbaci, S., and Tapada-Berteli, T. (2012). Social inequality and urban regeneration in Barcelona city centre: reconsidering success. *European Urban and Regional Studies* 19 (3), 287-311.

Ashworth, G., Page, S. J. (2011). Urban tourism research: Recent progress and current paradoxes. *Tourism management* 32(1), 1-15.

Benner, M. (2019). From overtourism to sustainability: A research agenda for qualitative tourism development in the Adriatic. *Working Paper* Heidelberg University. 92213. URL: https://mpra.ub.uni-muenchen.de/92213/ (accessed on 23.6. 2020).

Bellini, N, Go, F.M, Pasquinelli, C. (2016). Urban tourism and city development: Notes for an integrated policy agenda. In *Tourism in the City: Towards an Integrative Agenda on Urban Tourism* (pp. 333–339). doi:10.1007/978-3-319-26877-4_24

Blázquez-Salom, M., Blanco-Romero, A., Guall, J., Murray, I. (2019). Tourist gentrification of retail shops in Palma (Majorca). In: Milano, C., Cheer, J.M., and Novelli, M. (eds.): *Excesses, Discontents and Measures in Travel and Tourism*. London: CABI Publishing, 39–69.

Bol-Esteve, A.J., Arnandis-i-Agramunt, R. (2020). Touristificacion in the Central Market of Valencia: Fact or Fiction? In Ribeiro Almeida *et al.* (eds) *Handbook of Research on the Impacts, Challenges, and Policy Responses to Overtourism*. Hershey: IGI Global Publishing, 156-175

Botti, L., Peypoch, N. (2013). Multi-criteria ELECTRE method and destination competitiveness. *Tourism Management Perspectives* 6, 108-113.

Canestrelli, E., Costa, P. (1991). Tourist carrying capacity: A fuzzy approach. *Annals of Tourism Research* 18 (2), 295-311

Carballo, R., León, C., Carballo, M. (2019). Fighting overtourism in Lanzarote (Spain). *Worldwide Hospitality and Tourism Themes* 11(5), 506-515

Chollet, F., Allaire, J.J. (2018). *Deep Learning with R*. New York: Manning Publications Co.

Crouch, G.I., Ritchie, J.R. (1999). Tourism, competitiveness, and social prosperity. *Journal of Business Research* 44, 137–152.

Doxey, G. (1975). A causation theory of visitor-resident irritants, methodology and research inferences. The impact of tourism. (pp.195-198). Paper presented at VIth Annual Conference of the Travel Research Association. San Diego, USA.

Dwyer, L., Forsyth, P., Rao, P. (2000). The price competitiveness of travel and tourism: a comparison of 19 destinations. *Tourism Management* 21(1), 9–22.

Dwyer, L., Kim, C. (2003). Destination competitiveness: determinants and indicators. *Current Issues in Tourism* 6 (5), 369–414.

Dodds, R., Butler, R.W. (2019) *Overtourism.* Berlin: De Gruyter

Escudero Gómez, L.A. (2019). Residents' Opinions and Perceptions of Tourism Development in the Historic City of Toledo, Spain. *Sustainability* 11(14), 3854-3878.

Exceltur (2013). UrbanTUR 2012: *Monitor de Competitividad Turística de los destinos urbanos españoles*. URL: https://www.exceltur.org/urbantur/_(accessed on 19 January 2019)

Exceltur (2017). UrbanTUR 2016: *Monitor de Competitividad Turística de los destinos urbanos españoles*. URL: https://www.exceltur.org/urbantur/_(Accessed on 19.01.2019)

Exceltur (2019). MonitTUR: Monitor de Competitividad Turística relativa de las Comunidades Autónomas Españolas. Edición 2018.Madrid.

Favre, C. (2017). The Small2Mighty tourism academy. Growing business to grow women as a transformative strategy for emerging destinations. *Worldwide Hospitality and Tourism Themes* 9(5), 555-563

Fletcher, R., Murray Mas, I. Blanco-Romero, A., Macià Blázquez-Salom (2019) Tourism and degrowth: an emerging agenda for research and praxis, *Journal of Sustainable Tourism* 27:12, 1745-1763

Ghasemi, V. (2019) Residents' apathy and its influence on tourism development. Doctoral Dissertation Summary. *European Journal of Tourism Research* 22. 193-197

Getz, D. (1994). Residents' attitudes towards tourism. *Tourism Management* 15 (4), 247-258.

Goodwin, H (2017). The Challenge of Overtourism. Working Paper 4 Responsible Tourism Partnership: URL: https://haroldgoodwin.info/pubs/RTP\T1\textquoterightWP4Overtourism01\T1\textquoteright2017.pdf (Accessed on 19.01.2019).

Guttentag, D. (2015). Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector. *Current Issues in Tourism* 18 (12), 1192-1217.

Hall, C. M., Gössling, S., Scott, D. (2015). The evolution of sustainable development and sustainable tourism. In Hall, C. M., Gössling, S. & Scott, D. (Eds.). *The Routledge handbook of tourism and sustainability*. New York: Routledge, 15-25

Han, J., Kamber, M., Pei, J. (2012). *Data Mining. Concepts and Techniques*. Waltham. United States: Morgan Kaufmann Publishers.

Hassan, S. S. (2000). Determinants of market competitiveness in an environmentally sustainable tourism industry. *Journal of Travel Research* 38, 239–245.

Huete, R., Mantecón, A. (2018). El auge de la turismofobia ¿hipótesis de investigación o ruido ideológico? *Pasos. Revista de Turismo y Patrimonio Cultural* 16 (1), 9-19

Hunter, C. (2002). Aspects of the sustainable tourism debate from a natural resources perspective. In Harris, R., Griffin, T. & Williams, P. (Eds), *Sustainable tourism: A global perspective*. Oxford: Butterworth-Heinemann, 3-23

INE (2016) Contabilidad Regional de España (Spanish Regional Accounts). URL: www.ine.es. (Accessed on 19.01.2019)

INE (2017a). Encuesta de Ocupación Hotelera. Hotels occupancy survey, price index and profitability indicators. URL: www.ine.es (Accessed on 19.01.2019)

INE (2017b). Cifras oficiales de población municipal. Revisión del padrón municipal. Official population figures from Spanish municipalities: Revision of the Municipal Register. URL: www.ine.es (Accessed on 19.01.2019)

Koens, K., Postma, A., Papp, B. (2018). Is Overtourism Overused? Understanding the Impact of Tourism in a City Context. *Sustainability* 10 (4348), 1-15

Kuhn, M. (2020). caret: Classification and Regression Training. R package version 6.0-86. URL: https://CRAN.R-project.org/package=caret

Kuhn, M., Johnson, K. (2013). *Applied Predictive Modeling*. New York: Springer

Long, P.H., K. Kayat (2011) Residents' perceptions of tourism impact and their support for tourism development: the case study of Cuc Phuong National Park, Ninh Binh province, Vietnam. *European Journal of Tourism Research* 4(2), 123-146

Mansilla, J., Milano, C. (2018). Introducción a la ciudad de vacaciones. Apuntes sobre turismo y malestar social en Barcelona. In: Milano, C., Mansilla, J. *Ciudades de Vacaciones. Conflictos urbanos en espacios turísticos*. Barcelona: Pol·len edicions.

Martin, C. J. (2016). The sharing economy: a pathway to sustainability or a nightmarish form of neoliberal capitalism? *Ecological Economics* 121, 149-159.

Martín Martín, J. M., Guaita Martínez, J. M., Salinas Fernández, J. A. (2018). An analysis of the factors behind the citizen's attitude of rejection towards tourism in a context of overtourism and economic dependence on this activity. *Sustainability* 10(8) 2851, 1-18.

McKinsey&Company (2017). *Coping with success. Management overcrowding in tourism destinations. (Travel, Transport & Logistics)*. New York: World Travel &Tourism Council.

Meyer, D., Dimitriadou, E., Hornik, K.., Weingessel A., Leisch, F. (2019). e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.7-0.1. URL: https://CRAN.R-project.org/package=e1071

Michalkó, G., Bakucz, M., Rátz, T. (2013). The relationship between tourism and residents' quality of life: A case study of Harkány, Hungary. *European Journal of Tourism Research* 6(2), 154-169.

Milano, C. (2018). Overtourism, malestar social y turismofobia. Un debate controvertido. *Pasos. Revista de Turismo y Patrimonio Cultural* 16 (3), 551-564

Milano, C., Cheer, J. M., Novelli, M. (Eds.). (2019). *Overtourism: Excesses, discontents and measures in travel and tourism*. Wallingford: CABI.

Moreno, A. and Pardo, D. (2020). PGIS unidos contra los tiburones del turismo. *Crítica Urbana* 10, 34-37

Moreno-Izquierdo, L., Ramón-Rodríguez, A. B., Such-Devesa, M. J., Perles-Ribes, J. F. (2019). Tourist environment and online reputation as a generator of added value in the sharing economy: The case of Airbnb in urban and sun-and-beach holiday destinations. *Journal of Destination Marketing & Management* 11, 53-66.

Muler-González, V., Coromina, LL. and Galí, N. (2018). Overoturism: residents' perceptions of tourism impact as an indicator of resident social carrying capacity – case study of a Spanish heritage town. *Tourism Review*, DOI: 10.1108/TR-08-2017-0138.

Namberger, P., Jackish, S., Schmude, J. and Karl, M. (2019). Overcrowding, Overtourism and Local Level Disturbance: How Much Can Munich Handle?. *Tourism Planing & Development* 16(4), 452-472.

Neuts, B., Nijkamp, P. (2012). Tourist crowding perception and acceptability in cities: An applied modelling study on Bruges. *Annals of Tourism Research* 39 (4), 2133-2153.

Nunkoo, R., Ramkissoon, H. (2012). Power, trust, social exchange and community support. *Annals of Tourism Research* 39 (2), 997-1023.

Opillard, F. (2016). From San Francisco's 'Tech Boom 2.0'to Valparaíso's UNESCO World Heritage Site: resistance to tourism gentrification from a comparative political perspective. In Colomb, C., Novy, J. (2016). *Protest and resistance in the tourist city*. London: Routledge, 143-165.

Oklevik, O.Gössling, S. Hall, M. C. Steen Jacobsen, J. K. Grøtte, I.P., McCabe S. (2019) Overtourism, optimisation, and destination performance indicators: a case study of activities in Fjord Norway, *Journal of Sustainable Tourism* 27 (12), 1804-1824.

Peeters, P, Gössling, S., Klijs, J., Milano, C. Novelli, M., Dijkmans, C., Eijgelaar, E., Hartman, S., Heslinga, J., Issac, R., Mitas, O., Moretti, S., Nawijn, J., Papp, B., Postma, A. (2018). Research for TRAN Committee - *Overtourism: impact and possible policy responses*. Brussels: European Parliament, Policy Department for Structural and Cohesion Policies.

Perles, J.F., Ivars, J.A. (2019). Smart sustainability: a new perspective in the sustainable tourism debate. *Journal of Regional Research* 42, 151-170

Perles, J.F, Ramón, A.B, Vera, J.F., Ivars, J.A. (2018). The end of growth in residential tourism destinations: steady state or sustainable development? The case of Calpe. *Current Issues in Tourism* 21 (12), 1355-1385

Perles, J., Ramón, A., Moreno, L., Such, M.J. (2020) Tourism competitiveness and the well-being of residents: a debate on registered and non-registered accommodation establishments. *European Journal of Tourism Research* 24, 2406.

Perles, J.F, Ramón, A.B., Sevilla, M. (2014). La cuota de mercado como indicador de competitividad en los destinos turísticos: sentido y limitaciones. *Cuadernos de Turismo* 34, 265-285.

Perles, J.F., Ramón, A.B., Moreno, L., Sevilla, M. (2016). Economic crises and market performance—A machine learning approach. *Tourism Economics* 23 (3), 692-696

Perkumienė, D., Pranskūnienė, R. (2019). Overtourism: Between the right to travel and residents' rights. *Sustainability* 11 (7), 2138.

Plaza, A. (2018). Paradoja del turismo español: regiones cada vez más ricas con habitantes más pobres at URL: https://www.elconfidencial.com/economia/2018-10-04/turismo-espana-regiones-ricas-habitantes-pobres_1625210/ (Accessed 18.02.2020).

Poop, M. (2012). Positive and negative urban tourist crowding: Florence, Italy. *Tourism Geographies* 14 (1), 50-72

Porter, M. E. (1990). *The Competitive Advantage of Nations*. New York: Free Press.

Postma, A., Schmuecker, D. (2017). Understanding and overcoming negative impacts in city destinations: conceptual model and strategic framework. *Journal of Tourism Futures* 3(2), 144-156.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Ram, Y.; Hall, C.M. (2018). Walking tourism in cities: Introducing the special issue. *International Journal of Tourism Cities* 4, 281–284

Renda, A. I., da Costa Mendes, J., do Valle, P. O. (2014). The destination is where I live! Residents' perception of tourism impacts. *Journal of Spatial and Organizational Dynamics* 2(1), 72-88.

Ritchie, J. B., Crouch, G. I. (2003). *The competitive destination: A sustainable tourism perspective*. Wallingford: CABI.

Santos, J. A., Sequera, J. (2018). Expansión de la ciudad turística y nuevas resistencias: el caso de Airbnb en Madrid. Empiria: *Revista de metodología de ciencias sociales* 41, 15-32.

Seraphin, H., Sheeran, P., Pilato, M. (2018). Overtourism and the fall of Venice as a destination. *Journal of Destination Marketing and Management* 9, 374-376.

Sharpley, R. (2014). Host Perceptions of Tourism: A Review of the Research. *Tourism Management* 42, 37-49.

Silva, F. P. D., Brandão, F., Sousa, B. (2019). Towards socially sustainable tourism in cities: Local community perceptions and development guidelines. *Enlightening Tourism: A Pathmaking Journal (ET)* 9(2), 168-198

Spence, A., Hazard, H. (1988). *International competitiveness*. Cambridge, MA: Ballinger Publishing Company

Stanchev, R. (2018). The Most Affected European Destinations by Over-Tourism; Working Paper at University of Illes Balears: http://hdl.handle.net/11201/148140.

UNWTO; Centre of Expertise Leisure, Tourism & Hospitality; NHTV Breda University of Applied Sciences & NHL Stenden University of Applied Sciences (2018). *'Overtourism'?-Understanding and Managing Urban Tourism Growth beyond Perceptions*. Executive Summary. Madrid: UNWTO

Vera-Rebollo, J. F. V., Ivars-Baidal, J. A. I. (2003). Measuring Sustainability in a Mass Tourist Destination: Pressures, Perceptions and Policy Responses in Torrevieja, Spain, *Journal of Sustainable Tourism* 11(2-3), 181-203.

Vivés-Miró, S., Rullan, O. (2017). Desposesión de vivienda por turistización? Revalorización y desplazamientos en el Centro Histórico de Palma (Mallorca), *Revista de Geografía Norte Grande* 67, 53-71

Volgger, M.; Huang, S.S. (2019). Scoping irresponsible behaviour in hospitality and tourism: Widening the perspective of CSR. *International Journal of Contemporary Hospitality Management* 31(6), 2526-2543.

Wang, S (2016). Roles of place identity distinctiveness and continuity on resident attitude toward tourism. *European Journal of Tourism Research* 13, 58-68.

World Economic Forum WEF (2019). The Travel and Tourism Competitiveness Report 2019. Travel and Tourism at a Tipping Point. Geneva: World Economic Forum.

Zhang, Y., Ling, Ch. (2018). A strategy to apply machine learning to small datasets in materials science. *Npj Computational Materials* 4 (25), 1-8.

Ziliak, T., Mc.Closkey D.N. (2008). *The Cult of Statistical Significance. How the Sandard Error Cost Us Jobs, Justice, and Lives*. Michigan: The University of Michigan Press.

Received: 14/04/2020
Accepted: 22/07/2020
Coordinating editor: Stanislav Ivanov