scientific reports

Check for updates

OPEN

Online reviews explain differences in coastal and inland tourists' satisfaction

Giulia Contu, Cinzia Dessí, Carla Massidda & Marco Ortu

To achieve a comprehensive understanding of the factors influencing tourist satisfaction, there has recently been an increasing interest in the information provided by online reviews. In this regard, a rather unexplored issue concerns the causal relationship between topics and emotions expressed by consumers in the written text and their overall quality assessment given through a rating system. This study aims to contribute to filling this gap by investigating whether there are differences between coastal and inland Sardinian hotels in the topics and emotions expressed by online reviewers and how both affect customer satisfaction. To this end, we apply the new TOBIAS method which models the impact of topics, moods, and emotions contained in reviews on the level of satisfaction expressed by customers through the number of stars. The novelty of this method is that it combines natural language processing and causal inference to explain the customer's overall quality rating.

The Travel and Tourism (T&T) sector stands as a pivotal source of revenue and employment on a global scale, evincing consistent expansion over successive decades. Nonetheless, it confronts formidable international competition, compelling extant destinations to sustain their market standing through persistent endeavors or seize novel frontiers of global demand. In this scenario, the concept of tourist satisfaction is undoubtedly among those that have attracted the most attention in recent decades[1]. The interest in understanding tourist satisfaction is related to the influence of tourist happiness on the development of a destination and the profitability of private businesses. Higher satisfaction can stimulate tourist expenditure, repeat visits, positive recommendations, and reputation enhancement. Therefore, measuring tourist satisfaction and its determinants becomes crucial for policymakers and managers interested in improving the tourism supply of their destination[2]. Several empirical studies focus on these issues following both qualitative and quantitative approaches. Among these, increasing interest has recently emerged in measuring tourist satisfaction directly or indirectly from online reviews and star ratings[3]. In this regard, while today it is commonly acknowledged that feelings and emotions in online reviews influence customer satisfaction, the empirical relationship between review content and overall quality rating (i.e. number of stars) remains a rather unexplained question[4].

This study aims to contribute to filling this gap by investigating information contained in online tourist reviews, in the form of text and scores, for hotels located in Sardinia, one of the two major Italian islands. Sardinia Island presents a unique and compelling case study for examining tourist satisfaction through online reviews. As one of the major tourist destinations in the Mediterranean, Sardinia offers a diverse range of tourism experiences, from its renowned coastal areas to its lesser-explored inland regions. This dichotomy allows for an in-depth analysis of how different geographical contexts within the same overall destination influence tourist satisfaction. Moreover, Sardinia's economy heavily relies on tourism, making the insights derived from this single case study particularly relevant for local stakeholders, policymakers, and businesses aiming to enhance their service offerings. The island's distinct blend of natural beauty, cultural heritage, and varied tourism infrastructure provides a rich dataset for analyzing the nuanced impacts of different tourism environments on customer satisfaction. Thus, focusing on Sardinia is an appropriate approach for a case study[5] and enables us to explore the broader implications of our findings while maintaining a clear and specific context. Specifically, we analyzed more than 99 thousand reviews regarding 614 hotels, from May 2021 to April 2023, retrieved Booking.com platforms. A central issue of the analysis is the distinction between the hotels located in the inland territories of Sardinia and those located in the coastal areas. This is an interesting dichotomy to observe as, while the island is a well-known sun-and-sand destination, private and public interest in expanding inland tourism has grown significantly. However, This intention is consistent with the growing interest that the international tourist demand has shown in new forms of experiential tourism linked to local customs and traditions. In this context, the inland areas of Sardinia, with their communities, can represent a crucial asset to attract new segments of

Department of Business and Economics Sciences, University of Cagliari, Viale Fra Ignazio 17, Cagliari, Italy. ✉email: marco.ortu@unica.it

tourist demand and thus extend the tourism season in many internal areas of Sardinia, the quality and quantity of the services offered struggle to fully satisfy the standards required to attract increasingly demanding tourists, mainly because the decades of experience of the operators dealing with seaside tourism are lacking. In this scenario, online reviews represent a source of invaluable information. The main idea is to use natural language processing to understand whether there are differences between coastal and inland hotels in the topics used by online reviewers and how these topics impact on their satisfaction expressed through a final rating. In more detail, the proposed investigation is developed around the following main research questions:

RQ1 Do customer online reviews explain customer satisfaction (rating) provided for hotels in Sardinia?
RQ2 Which topics in online reviews impact positively or negatively on customer satisfaction (rating)?
RQ3 Are the topics that emerged from the analysis consistent between coastal and inland hotels?

To retrieve and process data from the web, we apply an adapted version of a new method recently proposed by[6].

This method's novelty relies on combining natural language processing and causal inference to explain customers' assessment of a phenomenon. This methodology, denoted as TOBIAS (TOpic modeling Based Index Assessment through Sentiment), is designed to quantify the influence of sentiments and topics within customers' comments describing a phenomenon over the level of satisfaction expressed by customers through the final rating assigned to the experience.

TOBIAS model is built by combining different techniques and methodologies. Firstly, it uses Sentiment Analysis to identify the emotional state expressed in textual form such as sentiments, emotions, and moods. Then, through Topic Modeling, it finds the main relevant topics inside customer online review comments. Finally, by Partial Least Square Path Modeling (PLS-PM henceforth), TOBIAS estimates the impact of topics on an overall rating that summarizes the quality assessment perceived by customers of an analyzed phenomenon.

The remainder of the paper unfolds as follows. "The general context" section delineates the contextual framework within which the analysis is situated. "Relevant literature and research purpose" section engages with pertinent literature and formulates research hypotheses. "Methodology" section outlines the methodology employed. "Methodology" section deliberates on the findings. Finally, "Concluding remarks, implications and future developments" section offers concluding remarks.

The general context

Sardinia is an island in the center of the Mediterranean Sea renowned for its fabulous beaches, pristine coastline, stunning landscapes, rich history, and vibrant culture. It should not be surprising, therefore, that tourism represents one of the main drivers of this Region's economy. However, it is well known, tourism flows are concentrated in the summer season (Figs. 1 and 2) and mainly concern coastal municipalities (Figs. 3 and 4).

Yet, Sardinia offers more than just its sea and beaches. Thanks to a rich endowment of environmental and cultural heritage, the island holds promising opportunities for tourism development beyond its coastal regions, especially within its inland areas. However, despite a growing interest, tourism during the off-peak seasons, beyond the summer months, struggles to gain momentum (Fig. 2).

Sardinia's untapped tourism potential presents a crucial opportunity to uplift communities facing economic and demographic challenges. The island grapples with depopulation, mirroring Italy's rural trends. Migration to

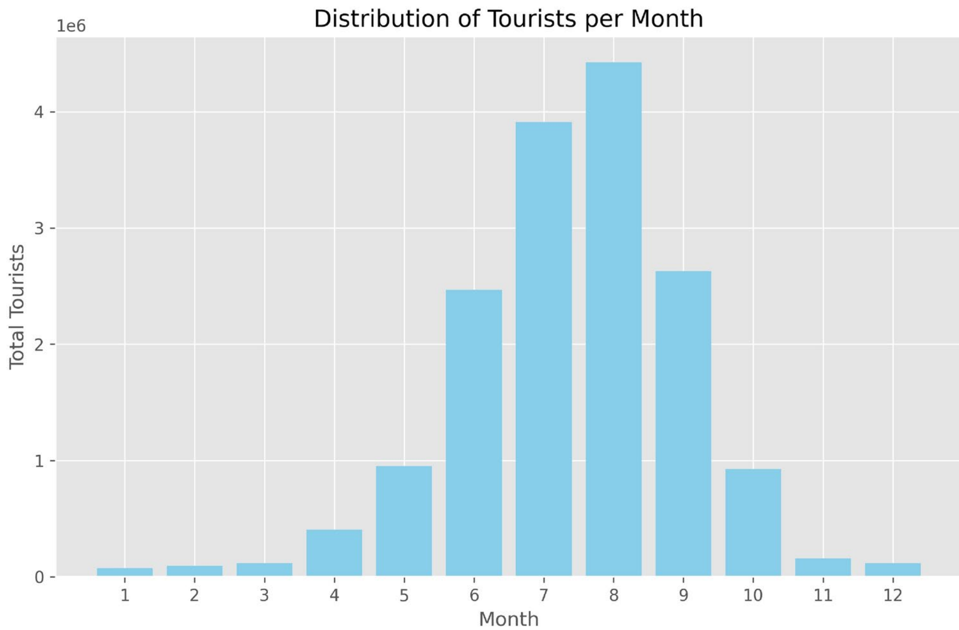

Fig. 1. Tourist nights by month. *Source* Our own elaboration on Regional Statistics Service 2022.

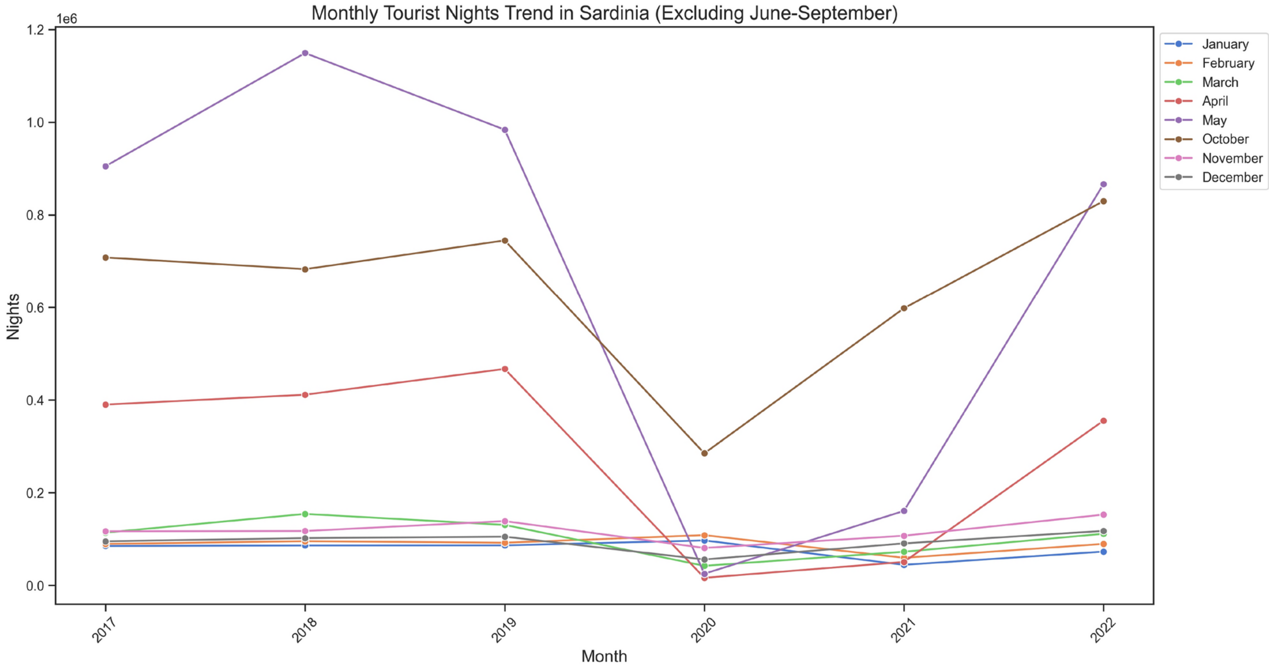

Fig. 2. Monthly nights trend in Sardinia by year (excluding June–September). *Source* Our own elaboration on Regional Statistics Service 2022.

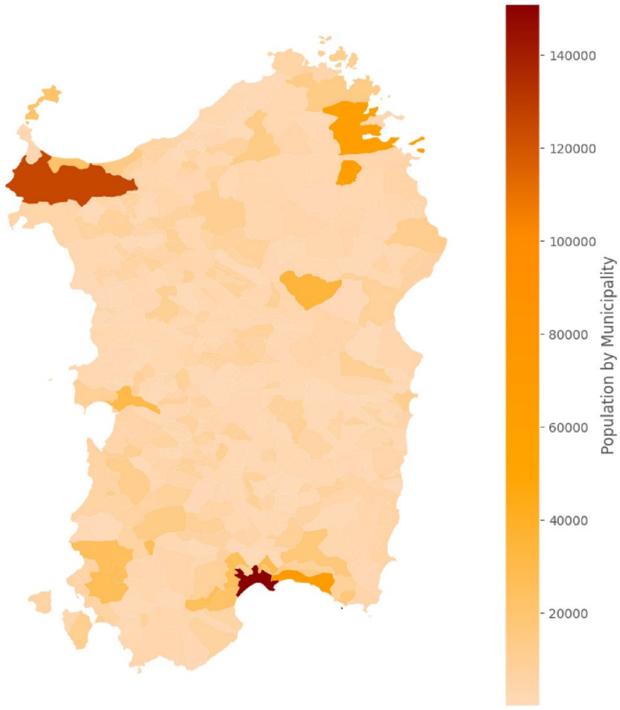

Fig. 3. Population by municipality. *Source* Our own elaboration on Regional Statistics Service 2022, the map is generated using an ad-hoc Python script.

urban areas predominantly situated along coastal areas (Fig. 3) strains resources, exacerbating aging population issues and necessitating improvements in healthcare, welfare, and social services.

This uneven demographic distribution is also reflected in the island's economy. Sardinia's economy is marked by stark disparities between its coastal and inland regions, exacerbated by the island's overall economic lag with respect the European regions. Despite a GDP exceeding that of Southern Italy by 6%, it still falls below

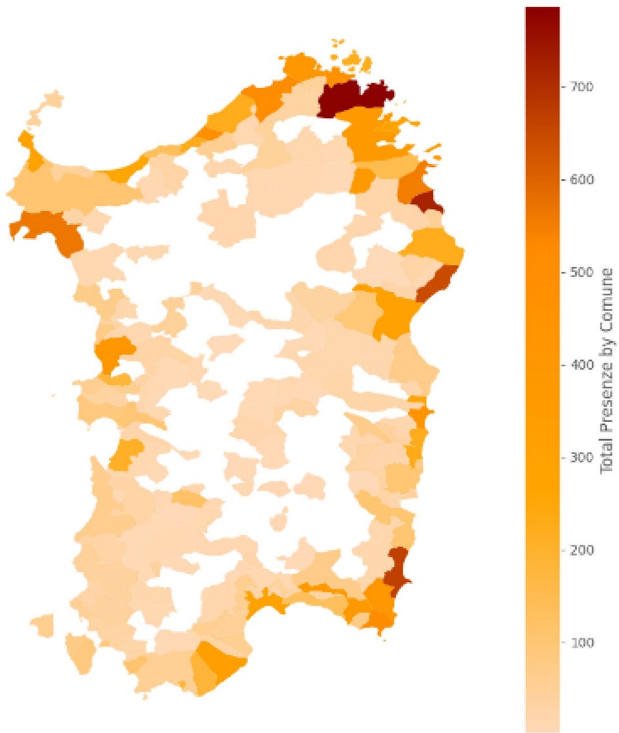

Fig. 4. Nights by municipality. *Source* Our own elaboration on Regional Statistics Service 2022, the map is generated using an ad-hoc Python script.

national and Central-Northern averages. Efforts by authorities and tourism operators aim to diversify offerings and improve tourism services, but challenges persist in meeting evolving tourist demands, especially in inland areas lacking seaside tourism experience.

Relevant literature and research purpose

The digital revolution and online platforms are reshaping tourism, altering customer behavior throughout their travel experiences[7]. With mobile devices and social media, travelers share opinions, photos, and reviews, influencing a wide audience. The impact of electronic word of mouth (e-WOM) is profound, as it reaches vast online audiences with enduring accessibility. Unlike traditional word of mouth, e-WOM offers global reach and indefinite retention, making it highly influential and reliable. As a consequence, consumers heavily rely on e-WOM when making travel purchases, driven by the significant investment and desire to minimize search efforts[8]. At the same time, scholars, decision-makers, and practitioners are increasingly interested in the critical role that e-WOM plays in shaping tourists' decision-making process[9–12]. It is therefore not surprising to see the proliferation of empirical studies interested in measuring tourist satisfaction starting from online reviews. A recent trend is to try to connect two forms of judgment that very often appear together, namely text and star rating[3]. With the stars, reviewers declare their general satisfaction with the products or services consumed, while with comments they express their emotional state (such as sentiment, emotions, and moods), describe their experiences, and provide recommendations. However, while today it is common opinion that there is a relationship between text and score, the causal relationship between these two components of the online review is a rather unexplained topic[4].

This literature gap offers ample room for investigation concerning the several elements composing a tourism product, but it matters especially for the hospitality industry where online customer rating and reviewing play a crucial role[13]. There are various examples of sentiment analysis regarding hospitality[14–17], and significant works in topic modeling[14,18]. In particular several recent studies have made significant contributions to the understanding of sentiment analysis and its application to consumer reviews in the tourism and hospitality industry. In[19], authors conducted a comparative assessment of sentiment analysis and star ratings for consumer reviews, demonstrating that sentiment analysis can offer deeper insights into customer perceptions beyond numerical ratings. In[20], it is explored the reasons why customers do not revisit in the tourism and hospitality industry, highlighting the critical role of negative sentiments in influencing customer retention.[21] examined tourist satisfaction with hotel services in Nainital, providing empirical evidence on the impact of service quality on overall satisfaction through sentiment analysis. Lastly,[22] utilized structural topic modeling to perform a bibliometric study of sentiment analysis literature, revealing key trends and thematic patterns that underscore the importance of integrating sentiment analysis in understanding customer feedback. These studies collectively reinforce the relevance and applicability of sentiment analysis in tourism research, supporting our investigation into the relationship between textual review components and overall customer satisfaction. There are also several attempts to combine these two approaches to investigate the underlying topics and sentimental polarity reflected in online reviews (see, for instance[23]), to associate sentiment with specific expressions (see, for instance[24,25]), and to reveal important information that is not visible to the viewer (see, for instance[26]). However, there are very few attempts to explore the consistency between sentiments expressed in online reviews and ratings dealing with the hospitality industry. To the best of our knowledge, only three examples are worth mentioning.

The first is the study[15], examines how customer sentiment polarity influences ratings for both premium and budget hotels in Goa. They conduct a regression analysis where the independent variable is customer sentiment polarity (defined as the ratio of positive to negative words in reviews), and the dependent variable is the customer rating. In this research, the Naïve Bayes algorithm was utilized to determine word polarity.

The second study provided by[27] examines the relationship between guests' sentiments and online ratings in the context of peer-to-peer accommodation on the base of 4602 reviews of San Francisco on the Airbnb platform. They found that positive (negative) sentiment was linked to high (low) ratings. Empirically they assess the role of sentiments in rating through a Tobit model where positive and negative sentiments interact with analytical thinking and authenticity. The model is completed with a set of control variables.

Lastly,[28] explore the connection between word clusters and customer satisfaction. Their study applies text mining to 8229 reviews from 25 hotels to identify key terms. A frequency analysis is conducted to extract the top 90 most frequent words, and a CONCOR analysis is used to form four distinct clusters. Subsequently, a regression model determines how these clusters influence the hotel customer satisfaction ratings.

While these studies have blazed new trails in research much work remains to be done and relevant limitations have been identified in previous works. For instance, some authors pointed out that analyzing words in isolation does not capture the overall sentiment of the reviewers[15], and that incorporating topic modeling with sentiment analysis is the way to better gauge the positive and negative connotations of the words[28]. Others suggest deeper analyses of the causal relationship between review content and star ratings[4].

These limitations altogether stimulate the present research proposal whose aim is to contribute to filling an important gap in the literature. The main idea is to bridge topic modeling with sentiment analysis to infer the interplay between single topics and consumer satisfaction. In more detail, our scope is to determine which topics have the highest influence on the rating and what the role of sentiments in the relationship between topic and rating. This innovative approach is developed to investigate the determinants of customer satisfaction for hotels located in Sardinia, an Italian island in the center of the Mediterranean Sea, and to highlight if there are differences in topics used by online reviewers between coastal and inland hotels. In summary, this research tries to answer the following questions.

RQ1 Do customer online reviews explain customer satisfaction (rating) provided for hotels in Sardinia?
RQ2 Which topics in online reviews impact positively or negatively on customer satisfaction (rating)?
RQ3 Are the topics that emerged from the analysis consistent between coastal and inland hotels?

Methodology

Theoretical and methodological underpinnings

As anticipated in the previous section, this study empirically investigates the determinants of customer satisfaction for hotels located in Sardinia (Italy) and highlights the differences between inland and coastal hotels. The investigation relies on information gained from online customer reviews in the form of text and scores (ratings).

Our goal is to investigate the determinants of customer satisfaction for hotels, with a leading case study of hotels located in Sardinia (Italy), and highlights the differences between inland and coastal hotels. Our approach is grounded in several theoretical frameworks that justify the assumption that the association between textual components of reviews and overall scores may vary across different geographical contexts. Expectation-Confirmation Theory (ECT)[29] posits that customers form expectations prior to a purchase, and their post-purchase satisfaction is determined by the degree to which these expectations are confirmed or disconfirmed by the actual experience. In the context of our study, we hypothesize that tourists' expectations for coastal and inland hotels differ significantly due to the distinct nature of these environments. Coastal areas, often more established in the tourism sector, might set higher expectations for quality and amenities, whereas inland areas, being relatively new to extensive tourism, might offer different experiences focused more on local culture and nature. These varying expectations are reflected in the online reviews and influence the satisfaction ratings provided by tourists. Cognitive Appraisal Theory[30] suggests that emotions result from an individual's cognitive evaluation of a situation or event. Applying this theory to our study, we argue that the emotional responses captured in online reviews are shaped by the tourists' appraisals of their hotel experiences. Coastal hotels, typically associated with traditional beach tourism, might evoke different emotions compared to inland hotels, which may offer unique cultural and natural experiences. These emotional responses are critical in determining the overall satisfaction ratings and are likely to vary between coastal and inland settings. The Two-Factor Theory[31], originally proposed in the context of workplace motivation, encounters several applications in tourism[32] and distinguishes between factors that cause satisfaction and those that cause dissatisfaction. In tourism, this theory can be applied to understand how different aspects of a hotel experience contribute to overall satisfaction or dissatisfaction. For instance, factors such as the quality of the room and the friendliness of the staff might be more crucial for coastal hotels, while the authenticity of the experience and engagement with local culture might be more significant for inland hotels. These factors, and the emotions they elicit, are expected to influence satisfaction ratings differently in coastal and inland areas. By integrating these theoretical frameworks, we hypothesize that the textual components of online reviews (topics and emotions) have different associations with overall satisfaction scores depending on whether the hotel is located in a coastal or inland area. This assumption is operationalized through our methodology, which combines natural language processing and causal inference to model these relationships. Here we use the term causal inference within the context of PLS-PM[33], where we hypnotize causal relationships among *latent constructs* and we measure causality in terms of predictive power. We utilize the TOBIAS method, which integrates Sentiment Analysis, Topic Modeling, and Partial Least Squares Path Modeling (PLS-PM). Sentiment Analysis captures the emotional content of the reviews, while Topic Modeling identifies the main themes discussed by the reviewers. PLS-PM then quantifies the relationships between these themes, the emotions they evoke, and the overall satisfaction ratings. Our methodological approach allows us to explore not only whether customer reviews explain satisfaction ratings but also how the impact of specific topics and emotions on satisfaction varies between coastal and inland hotels. This nuanced understanding can provide targeted insights for hotel managers and policymakers to enhance tourist satisfaction in different geographical contexts. In summary, the combination of ECT, Cognitive Appraisal Theory, and Two-Factor Theory provides a robust theoretical underpinning for our analysis. According to this theoretical perspective, satisfaction and dissatisfaction are perceived as distinct sentiments emanating from various and complex interactions between an external stimulus (such as a product or service) and the individual. Consequently, an individual may concurrently experience both satisfaction and dissatisfaction. Figure 5 highlights the theoretical framework that supports our analysis.

This methodological framework is applied to a case study focused on tourism data of Sardinia Island. The application of a single-case study approach is justified by the unique characteristics of Sardinia, which serves as a critical, unique, and revelatory case in understanding the relationship between online reviews and tourist satisfaction. According to[5], the single-case study is particularly suitable when the case exemplifies a unique or extreme circumstance that warrants in-depth exploration (Critical Case Testing). Sardinia's dual identity as both a coastal and inland tourist destination offers a distinct context that is not commonly replicated in other regions, making it an ideal subject for focused investigation (Unique or Extreme Case). Furthermore, the island's burgeoning interest in off-season and experiential tourism represents a revelatory opportunity to examine emergent trends that have been largely unexplored in other studies (Revelatory Case). By leveraging the comprehensive dataset of online reviews from Booking.com, which captures temporal trends across different tourist seasons, the study effectively utilizes a longitudinal aspect within a single-case framework (Longitudinal Case). Our unit of analysis is the individual review, which is crucial for understanding how specific comments and ratings reflect tourist satisfaction in different geographical areas within Sardinia. By analyzing each review, we can identify localized factors impacting satisfaction, thereby offering detailed insights into the coastal versus inland tourism experience. This granular approach aligns with the case study methodology, allowing for in-depth analysis within the chosen context, in particular, of the localized factors influencing tourist satisfaction, thereby providing insights that could inform tourism management strategies in similar but distinct contexts globally. Ultimately, focusing on a single, well-defined area also highlights the broader applicability of the findings to other regions with similar tourism dynamics.

To retrieve and process data from the web, we apply an adapted version of a method very recently proposed by[6]. This method, called TOBIAS, allows modeling the effects of the topics, and emotional state conveyed in the customers' comments describing a phenomenon, such as the perception of the quality of service, over the level of satisfaction expressed by customers through a rating system. TOBIAS exploits the textual content from reviews to infer and explain customer quality assessments and support quality assurance in improving the overall quality of services delivered to final customers.

This method's novelty relies on combining Natural Language Processing (NLP) techniques with causal inference through PLS-PM to construct a model that is both interpretable and robust. Compared with TOBIAS' original application, the present analysis introduces a topological change to split the Affectiveness latent variable into two new latent variables: i) Positiveness and ii) Negativeness. The introduction of these two variables enhances the model's interpretation, giving a direct and clearer view of the direction of negative and positive emotional states toward customer satisfaction.

TOBIAS method is developed through four main steps that describe the foundational components of its framework:

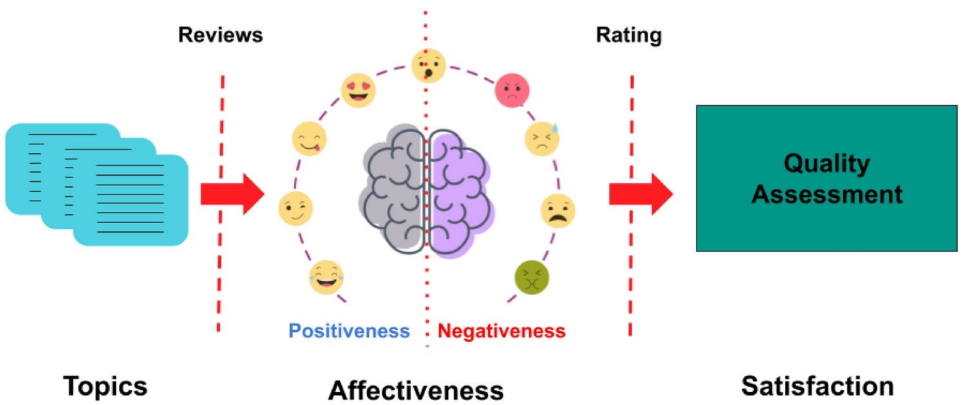

Fig. 5. Theoretical framework.

1. Emotional Features extraction through Sentiment Analysis techniques.
2. Topics assignment through Topic Modeling techniques.
3. Overall quality indexes definition.
4. PLS-PM model estimation.

Foundational components of TOBIAS

The first foundational component of TOBIAS is represented by Sentiment Analysis which can be defined as a process of obtaining and analyzing texts in order to identify attitudes and opinions expressed within[34]. More in detail, Sentiment Analysis attempts to differentiate between subjective and objective material by initially identifying subjectivity and objectivity in texts. It evaluates the polarity by allocating emotion ratings to text, which may be quantitative (limited to a particular range) or qualitative (positive/negative). Moreover, Sentiment Analysis investigates discrete emotions, which allows a more intricate examination by extracting emotions represented in human language, such as love and joy. We used a deep learning neural network to extract sentiment and discrete emotion from text[35]. This is a pre-trained model[36] on general human text, to reduce the subjectivity of this model we fine-tuned the model on a dataset of $\sim 20k$ hotel reviews with reviews labeled with positive and negative sentiment. On this dataset the pre-trained model obtained an accuracy of 0.71 and the fine-tuned model finally reached an accuracy of 0.94.

The second component is given by Topic Modeling which is an unsupervised learning method that recognizes underlying patterns and themes in a set of documents and facilitates their representation according to the frequency of the words that constitute them. We adopted a recently new technique: BERTopic[37]. The third component constitutes the Partial Least Squares Path Modeling (PLS-PM) estimation technique, serving as a least squares counterpart to Structural Equation Modeling reliant on maximum likelihood. A PLS-PM is delineated by two sets of linear equations, specifically the inner and outer models. The inner model elucidates the associations among latent variables, which are variables not directly observed. The inner model follows a causal chain structure, implying that the residuals are independent of each other, and no correlations exist between the residual terms associated with a particular endogenous latent variable and the latent variables that serve as its explanatory factors. The outer model concerns the relationships between latent and manifest directly observable variables. Figure 6 reports both models, where manifest variables are in pink and latent variables are in blue.

The inner model is represented in the following Eq. (1):

$$\xi = B\xi + \epsilon \tag{1}$$

In this context, ξ represents a vector of latent variables, B is the matrix of coefficients describing their interrelations, and ϵ denotes the residuals of the inner model. Specification of predictors reduces the Eqs. (1) and (2):

$$(\xi \mid \xi) = B\xi \tag{2}$$

The outer model delineates the interactions between latent and manifest variables, representing these manifest variables in two configurations: (i) reflective, and (ii) formative. In the reflective mode, causal paths originate from the latent variable (appearing on the right-hand side of Eq. 3) toward the manifest variables. Each manifest variable is consequently modeled as a linear function of its corresponding latent variables, inclusive of residuals. Conversely, the formative mode posits causal paths emanating from the manifest variables (on the left-hand side of Eq. 4) toward the latent variables. Thus, each latent variable is constructed as a linear function of its manifest variables, also incorporating residuals.

$$X_x = \Lambda_x \xi + \epsilon_x \tag{3}$$

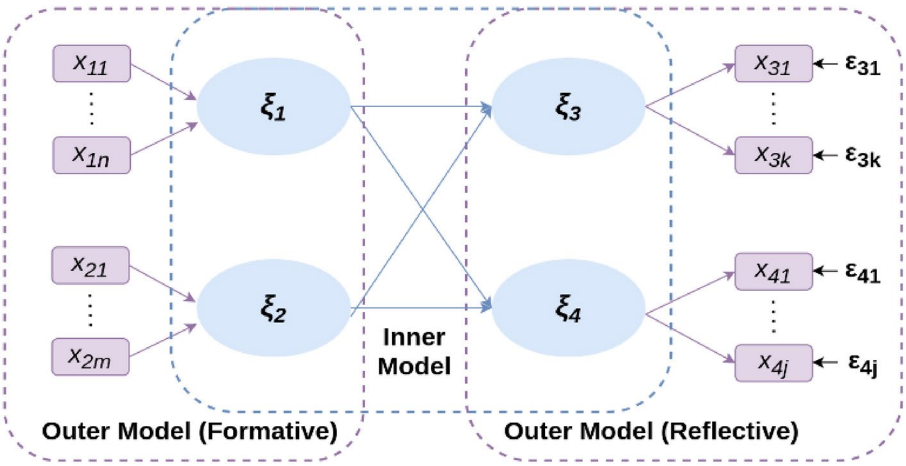

Fig. 6. Path model definition[6].

$$\xi = \Pi_x X_x + \epsilon_x \tag{4}$$

These equations under the predictors reduce to Eqs. 5 and 6.

$$(X_x \,|\, \xi) = \Lambda_x \xi \tag{5}$$

$$(\xi \,|\, X_x) = \Pi_x X_x \tag{6}$$

Fundamentally, PLS-PM algorithm operates through a series of regressions based on weight vectors. At convergence, the resultant weight vectors fulfill fixed point equations[38].

TOBIAS framework and algorithm

The following Fig. 7 adapts the conceptual framework in Fig. 6 to describe TOBIAS. Again, blue and pink nodes represent the latent and manifest variables respectively. In graphical representations, reflective manifest variables are depicted with an arrow pointing from the latent variable to the manifest variable. Conversely, formative manifest variables are illustrated with an arrow originating from the manifest variable and directed toward the latent variable. Yellow nodes represent Affective manifest variables. We denote with *H*, the number of Topics, with *K* the number of affect variables, and with *M* the number of Satisfaction indicators. What is proposed in Fig. 7 is an adapted version of the original TOBIAS model, which was first applied to university student's satisfaction[6]. As anticipated in "Foundational components of TOBIAS" sub-section, we introduce a topological change in the model to split the *Affectiveness* latent variable into two new latent variables: (i) *Positiveness* and (ii) *Negativeness*.

The inner model captures the causal link from *Affectiveness*, encompassing both *Positiveness* and *Negativeness*, to Satisfaction. The *H* latent variables, ξ_{θ_h} representing the topics, act as the primary catalysts that evoke *Affectiveness*. Moving on to the outer model, the *Satisfaction* block $\xi^\star$ reflects *M* manifest variables that represent the users' perceived overall quality of the services. The *Topics* blocks ξ_{θ_h} have a formative structure. Lastly, *Affectiveness* comprises a formative block that includes all measurable indicators of emotional state derived from the text, such as sentiment, emotion, and other related aspects. These manifest variables can be organized into *K* homogeneous sub-blocks *A*, each corresponding to a specific emotional state they characterize.

In the subsequent sections, we will delve into the specifics of the algorithm that executes the four primary phases of the TOBIAS model. Consider a set of *N* documents $D = \{d_1, \ldots, d_N\}$, where each document corresponds to a unique review *r* submitted by a user *u*, denoted by the paired element $(r, u) \in O$.

In this model, each document *d* is defined as a set of sentences $\{s_1, s_2, \ldots\}$, unordered and each marked by both emotional attributes and a pervasive topic. The aggregation of sentences across all *N* documents forms the set $S = \bigcup_{i=1}^{N} \{s \mid s \in d_i\}$, whose total number of elements, or cardinality, is $|S| = n_S$. For any given review, the set *S* can be divided into |*O*| groups $S_i = \{s \mid s \in S \wedge \gamma(s) = (u, r)_i\}$, where $\gamma : S \rightarrow O$ is a mapping function

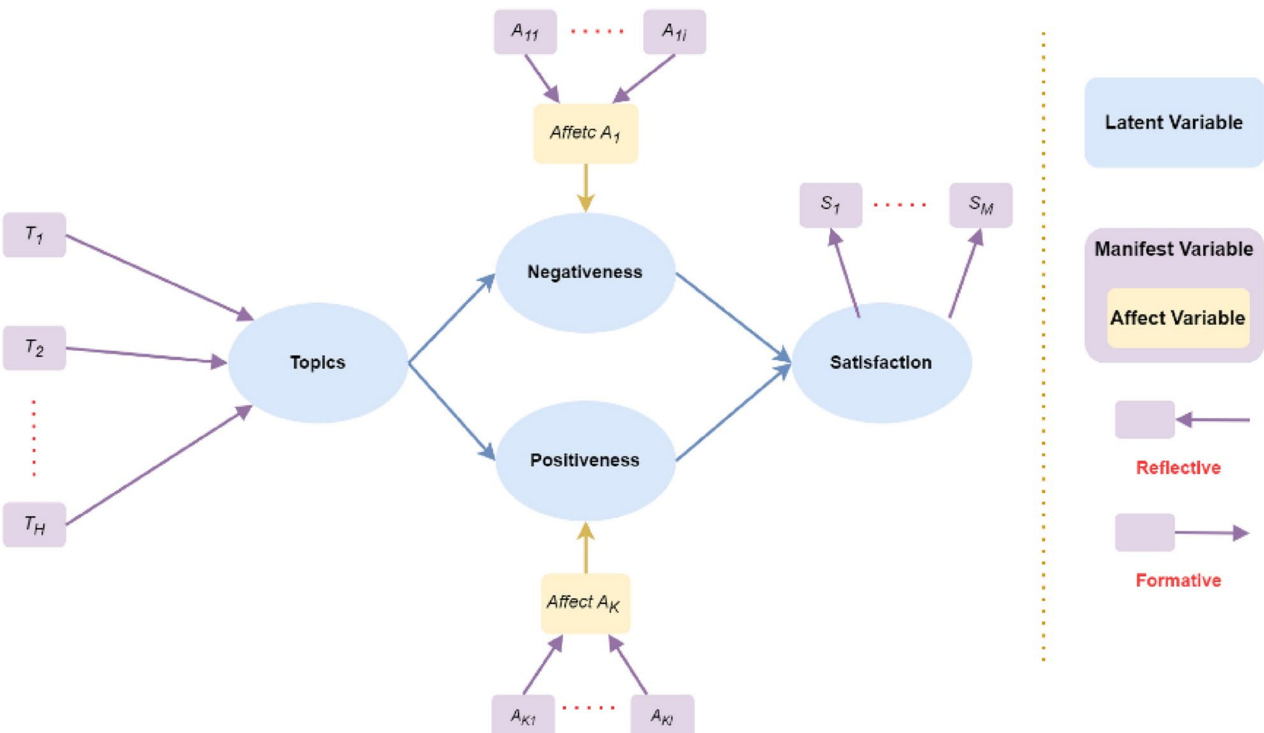

Fig. 7. TOBIAS framework.

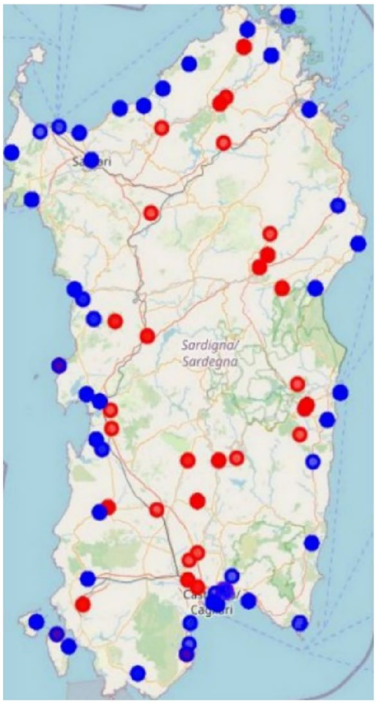

Fig. 8. Map of inland and coastal municipalities. The map is generated using an ad-hoc Python script.

Dataset statistics	Value
Reviews	99,505
Coastal reviews	76,130
Inland reviews	23,375
Hotels	614
Coastal hotels	396
Inland hotels	201

Table 1. Data description.

that assigns each sentence to its corresponding reviewer-user pair (*r*, *u*). The initial phase of the algorithm focuses on extracting emotional features. Consider a set of explicit emotional features *F* within the range [0, n_S], indicative of an underlying latent trait ξ_A, denoting Affectiveness. These features, *F*, are organized into *K* clusters A_k based on their emotional characteristics.

In the subsequent phase, called topic assignment, define *W* as the Bag-of-Words representing all words in *S*. Let $\Theta = \{\theta_1, \ldots, \theta_H\}$ represent all topics found within *S*. The topic modeling function $\varphi(\cdot)$ associates words from *W* to topics, with $\varphi : W \rightarrow \Theta$. Each topic θ_h is associated with a latent variable $\xi_{\theta h}$, represented by a unique manifest variable $F_{\theta h}$.

The third phase entails computing manifest indices that capture the latent overall quality of services $\xi_\star$ (Satisfaction) as perceived by end-users. Subsequently, values pertaining to the same pairs (*r*, *u*) are aggregated, adjusting the number of observations from n_S to the cardinality of *O*.

Data

The analysis is conducted using online reviews from Booking.com. Booking.com is one of the largest and most widely used online travel agencies globally, offering a comprehensive database of hotel reviews from a diverse range of travelers. This platform's extensive reach and popularity ensure that we have access to a large and varied sample of reviews, enhancing the robustness and relevance of our study. Additionally, Booking.com provides detailed reviews that include both numerical ratings and textual comments, allowing for a rich analysis of customer satisfaction from multiple dimensions[13]. We focus on the reviews regarding hotels located in Sardinia distinguishing between the hotels located in coastal and inland municipalities (Fig. 8).

Specifically, as detailed in Table 1, we analyze more than 99 thousand of reviews regarding 614 hotels, from May 2021 to April 2023. Out of 614 hotels, 396 are located in coastal municipalities corresponding to 76,130 reviews. To obtain a representative sample of all areas in Sardinia, we evenly sampled hotels all over the provinces. This sample is statistically significant, as the minimum sample size that guarantees a 95% confidence

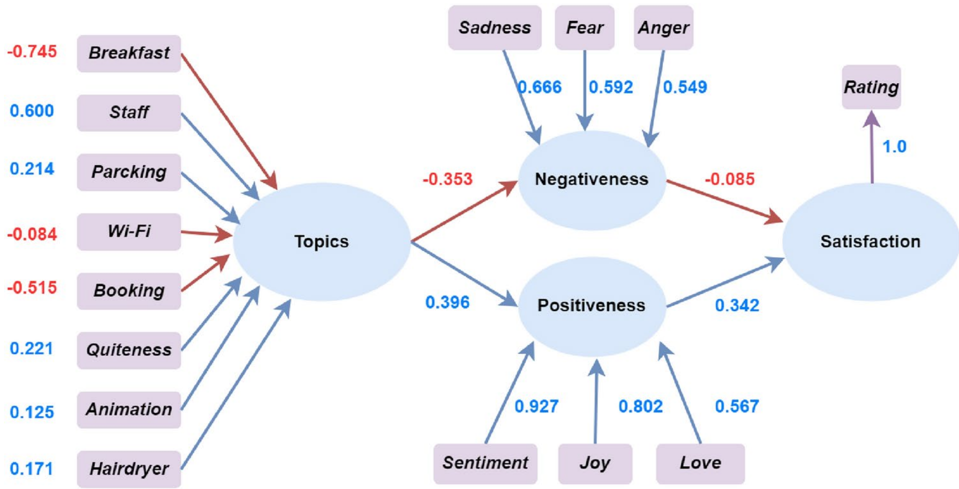

Fig. 9. TOBIAS results for coastal hotels. *Source* Our own elaboration on Regional Statistics Service 2022.

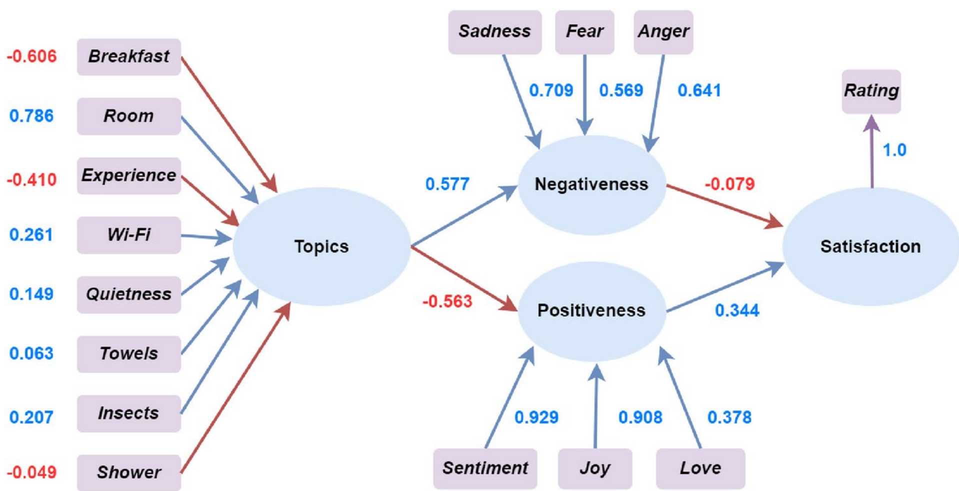

Fig. 10. TOBIAS results for inland hotels. Source: Our own elaboration on Regional Statistics Service 2022.

level with a 5% margin of error, assuming maximum variability is given by: $n = \frac{N \cdot (Z^2 \cdot p \cdot (1-p))}{(d^2 \cdot (N-1)) + (Z^2 \cdot p \cdot (1-p))}$, where N is the population size ($\sim$ 10,000), Z is the Z-score (1.96 for a 95% confidence level), p is the estimated proportion of an attribute that is present in the population (0.5 used for maximum variability), and d is the margin of error (0.05 for 5%). The calculation yields a minimum sample size of approximately ~ 370 hotels, therefore, with a sample of 614 hotels, we ensure the representativeness of the broader population of hotels.

Results

This section presents and discusses the main findings summarized in Figs. 9 and 10. Here, blue arrows represent positive relationships and red arrows represent negative relationships. The magnitude of each coefficient measures the strength of a specific relationship, while the sign indicates its direction. The general overview seems to indicate interesting differences based on hotel location.

Let us discuss the main findings starting from the inner model. This model demonstrates, through path coefficients obtained from linear regression, how the latent variable *Topics* impacts on *Affectiveness* (both *Positiveness* and *Negativeness*) and how the latter influences *Satisfaction*. Table 2 shows PLS-PM estimates reported in Figs. 9 and 10 with statistical testing. Here, bold values represent statistically significant constructs ($p < 0.05$), while the intercepts for all constructs are effectively zero (with corresponding *p*-values of 1.00), indicating that when the predictors are at their reference level, the constructs do not differ from their mean value. This is typical in PLS-PM where the primary interest lies in the relationships between constructs rather than the absolute values. Thus, based upon results in Table 2, we can assert that the latent variable *Topics* plays a significant role in explaining *Affectiveness* in both its positive and negative aspects. Moreover, different magnitudes and signs are estimated between hotels with different locations. For hotels located inland, an estimated coefficient 0.577 seems to indicate that an increase in *Topics* leads to an increase in negative sentiment, whereas the coefficient -0.563 indicates

Variable	Estimate	Std. error	t value	$Pr(>\|\|t\|\|)$
Inland				
Positiveness				
Intercept	− 0.00	0.061	− 0.00	1.00
Topics	− 0.563	0.062	− 10.021	***
Negativeness				
Intercept	− 0.00	0.064	− 0.00	1.00
Topics	0.577	0.063	10.391	***
Satisfaction				
Intercept	0.00	0.063	0.00	1.00
Positive	0.344	0.175	2.011	*
Negative	− 0.079	0.173	− 0.464	0.612
Coastal				
Positiveness				
Intercept	− 0.00	0.06	− 0.00	1.00
Topics	0.396	0.061	6.342	***
Negativeness				
Intercept	− 0.00	0.06	− 0.00	1.00
Topics	− 0.353	0.064	− 5.552	***
Satisfaction				
Intercept	0.00	0.061	0.00	1.00
Positive	0.342	0.161	2.13	*
Negative	− 0.085	0.165	− 0.534	0.632

Table 2. Structural assessment of TOBIAS model. Results of the significance of Affectiveness and Satisfaction Blocks. The statistical significance of blocks is indicated as follows: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

that a decrease in *Topics* results in an increase in positive sentiments. Therefore, both coefficients imply that the latent variable *Topics* tends to elicit stronger negative sentiments. On the contrary, for hotels situated along the coast an increase in the value of *Topics* leads to a decrease in negative sentiment (-0.353) and an enhancement of positive feelings (0.396). As for the impact of *Affectiveness* on *Satisfaction*, the result is homogeneous between the two groups of hotels divided by location. All together, these findings imply that the latent variables *Topics* and *Affectiveness* incorporated into the model accurately explain the latent variable *Satisfaction* and that the results for hotels in both land and coastal locations are comparable in terms of significance.

These initial findings answer the first research question (*RQ1*) the Topics and Affectiveness expressed in online reviews can explain customer Satisfaction.

Taking estimated coefficients in Table 2, the relationship linking *Satisfaction* with *Affectiveness* (*Positiveness* and *Negativeness*) and *Topics* is derived through the following formulas (corresponding to Eq. 2). The coefficients are computed using the ordinary least squares method and represent the change in the dependent variable in terms of standard deviations[39].

For inland and coastal locations:

$$\begin{aligned} Satisfaction_{\text{land}} &= -0.079 \cdot Negativeness + 0.344 \cdot Positiveness &= -0.079 \cdot (0.577 \cdot Topics) + 0.344 \cdot (-0.563 \cdot Topics) \\ Satisfaction_{\text{coast}} &= -0.085 \cdot Negativeness + 0.342 \cdot Positiveness &= -0.085 \cdot (-0.353 \cdot Topics) + 0.342 \cdot (0.396 \cdot Topics) \end{aligned} \tag{7}$$

Let us now move to the outer model whose outcomes show how manifest variables explain the latent variables in the formative model, and vice versa in the reflective model. The formative model's output allows for the definition of the latent variable *Topics* as a linear combination of the topics identified within the reviews. The linear model is defined as follows for both locations (corresponding to Eq. 5):

$$\begin{aligned} Topics_{\text{land}} &= -0.606 \cdot Breakfast + 0.786 \cdot Room + \cdots + (-0.049) \cdot Shower \\ Topics_{\text{coast}} &= -0.7449 \cdot Breakfast + 0.60 \cdot Staff + \cdots + 0.171 \cdot Hairdryer \end{aligned} \tag{8}$$

Each topic is incorporated into the model representing the probability of a review addressing each specific topic (given that a review is a mixture of topics).

Then, thanks to the bifurcation of *Affectiveness*, the model can estimate the impact of each topic on both *Negativeness* and *Positiveness*.

For inland locations:

Topic	Total effect on satisfaction
Breakfast	0.149
Room	**− 0.188**
Experience	0.098
Wi-Fi	**− 0.062**
Quietness	**− 0.036**
Towels	**− 0.015**
Insects	**− 0.050**
Shower	0.012

Table 3. Impact of each topic on satisfaction for land.

Topic	Total effect on satisfaction
Breakfast	**− 0.123**
Staff	0.099
Parking	0.035
Wi-Fi	**− 0.014**
Booking-Payment	**− 0.085**
Quietness	0.037
Animation	0.021
Hairdryer	0.028

Table 4. Impact of each topic on satisfaction for coast.

$$\begin{aligned} \text{Negativeness}_{\text{land}} &= +0.577 \cdot \text{Topics} &= +0.577 \cdot (-0.606 \cdot \text{Breakfast} + 0.786 \cdot \text{Room} + \cdots) \\ \text{Positiveness}_{\text{land}} &= -0.563 \cdot \text{Topics} &= -0.563 \cdot (-0.606 \cdot \text{Breakfast} + 0.786 \cdot \text{Room} + \cdots) \end{aligned} \tag{9}$$

For coastal locations:

$$\begin{aligned} \text{Negativeness}_{\text{coast}} &= -0.353 \cdot \text{Topics} &= -0.353 \cdot (-0.606 \cdot \text{Breakfast} + 0.786 \cdot \text{Room} + \cdots) \\ \text{Positiveness}_{\text{coast}} &= +0.396 \cdot \text{Topics} &= +0.396 \cdot (-0.606 \cdot \text{Breakfast} + 0.786 \cdot \text{Room} + \cdots) \end{aligned} \tag{10}$$

Furthermore, the output from the formative model allows for the definition of the latent variable *Affectiveness* as a linear combination of negative emotions, such as sadness, fear, and anger, and positive emotions, such as sentiment, joy, and love. The values of positive and negative *Affectiveness* can be calculated as follows:

$$\begin{aligned} \text{Negativeness}_{\text{land}} &= 0.709 \cdot \text{Sadness} + 0.569 \cdot \text{Fear} + 0.641 \cdot \text{Anger} \\ \text{Positiveness}_{\text{land}} &= 0.929 \cdot \text{Sentiment} + 0.908 \cdot \text{Joy} + 0.378 \cdot \text{Love} \end{aligned} \tag{11}$$

$$\begin{aligned} \text{Negativeness}_{\text{coast}} &= 0.666 \cdot \text{Sadness} + 0.592 \cdot \text{Fear} + 0.549 \cdot \text{Anger} \\ \text{Positiveness}_{\text{coast}} &= 0.927 \cdot \text{Sentiment} + 0.802 \cdot \text{Joy} + 0.567 \cdot \text{Love} \end{aligned} \tag{12}$$

Sadness, fear, anger, joy, and love are incorporated into the model as probabilities that reflect the likelihood of recording a specific emotion, while sentiment is measured on a scale that ranges from -1 to $+1$. The findings reveal similarities between inland and coastal hotels, with coefficients displaying similar values and signs. This indicates that emotions expressed in reviews are more significantly affected by the services offered rather than the hotel's location.

Finally, the outer model describes how the latent variable *Satisfaction* explains the manifest variable *Rating* in the reflective mode. As shown in Figs. 9 and 10, by construction there is a perfect correspondence between the two. In our analysis *Satisfaction* is measured only by the manifest variable *Rating*, thus they express the same phenomena.

By integrating the insights from both the inner and outer models, it becomes possible to understand how each manifest variable contributes to *Satisfaction*. For example, the manifest variable *Breakfast* exhibits a substantial negative coefficient for both inland (-0.606) and coastal hotels (-0.749). To determine the final impact on *Satisfaction*, it is necessary to multiply the coefficients, considering the various factors that define the outer and inner models. Table 3 and Table 4 show the final impacts of *Topics* on *Satisfaction*, and are obtained following the same multiplying rule. Upon examining the data in Table 3 and Table 4, it is noted that the final impact of breakfast on satisfaction amounts to 0.149 for hotels located inland and − 0.123 for those on the coast. To make the derivation of the coefficients clearer, we show below the calculation for inland hotels:

$$
\begin{aligned}
&(\text{Breakfast coefficient} \cdot \text{coefficient of topic} \cdot \text{coefficient of negativeness}) + \\
&(\text{Breakfast coefficient} \cdot \text{coefficient of topic} \cdot \text{coefficient of positiveness})
\end{aligned}
\tag{13}
$$

Substituting with the numbers, it is obtained the following formula:

$$
[(-0.606) \cdot 0.577 \cdot (-0.079)] + [(-0.606) \cdot (-0.563) \cdot 0.344] = 0.149 \tag{14}
$$

Tables 3 and 4 address the second research question (*RQ2*) topics highlighted in bold impact negatively on customer satisfaction while other topics impact positively.

Following the same approach, the final impacts of sentiments and emotions on satisfaction can be computed using the same multiplying rule of coefficients from the outer to the inner model. For instance, the impact of *Sentiment* on *Satisfaction* equals 0.319, and it is calculated as follows:

$$
(\textit{Sentiment}\ \text{coefficient} \cdot \textit{Positiveness}\ \text{coefficient}) = (0.929 \cdot 0.344) = 0.319 \tag{15}
$$

Upon analyzing Tables 3 and 4, it is evident that there are both similarities and differences in the impact of individual topics on the latent variable *Topics* for hotels located inland versus those on the coast. For example, topics such as breakfast, quietness, and Wi-Fi are mentioned in reviews, regardless of the hotel's location. However, as shown in the tables, the coefficients report different magnitudes depending on the location and, in the case of Wi-Fi, even with opposite signs. In contrast, topics like parking, animation, staff, and booking payments are only mentioned in reviews of coastal hotels, while inland hotels are more frequently associated with topics such as insects, the overall experience, and showers. Thus, specific topics are unique to certain areas. In terms of satisfaction, it also happens that the same topic reports opposite effect, such as breakfast which shows a positive sign for inland hotels and a negative for coastal ones.

These findings answer the third research question (*RQ3*) topics are significantly different depending on hotels' location in terms of both different topics discussed and the final impact on customer satisfaction.

In sum, we can conclude that the overall results of this study can provide robust answers to the three research questions leading the analysis. Compared to existing literature, with TOBIAS we have been able to provide a more comprehensive explanation of the impact of experience on customer satisfaction. At the same time, we confirm some previous findings. For instance, the fact that animation is positively associated with coastal hotels aligns with the results provided by[40] and[41]. We also confirm the central role of the topic breakfast, as reported in[42] and[28]. Furthermore, the negative impact of room quality in inland hotels aligns with results discussed by other researchers, such as[42–44]. We also confirm the significant role of the hotel's location which is consistent with previous studies indicating that perceptions of destinations and hotels are influenced by various factors, including culture, types of travelers, and location (e.g.,[45–47]). These results can be very useful considering the implications of tourist satisfaction on core tourism metrics. Several studies have explored the relationship between hotel ratings on platforms like Booking.com and core tourism metrics such as *revenues* and *tourist numbers*. In[13], it is examined how online consumer reviews and management responses on platforms like Booking.com impact hotel performance, including revenue. It highlights that higher ratings and positive reviews are correlated with increased revenue, as they enhance the hotel's reputation and attractiveness to potential guests. In[48], they analyze how online reviews affect hotel performance metrics such as occupancy rates, average daily rates (ADR), and revenue per available room (RevPAR). The study finds that higher ratings and more positive reviews significantly boost these key performance indicators. Moreover[49], discusses the influence of user-generated content, including reviews on Booking.com, on hotel performance. It demonstrates that higher online ratings lead to higher occupancy rates and increased revenue, emphasizing the importance of maintaining positive online reputations. Finally,[50,51] compares the predictive power of social media review ratings, including those from Booking.com, against traditional customer satisfaction metrics. It concludes that online ratings are strong predictors of hotel performance, particularly regarding revenue and tourist numbers. These studies collectively highlight the significant impact that online reviews and ratings have on hotel performance metrics such as revenue and tourist numbers. The insights that TOBIAS methodology can provide can be used to decrease negative aspects and increase positive aspects of reviews increasing both sentiment and ratings of hotels and as a consequence, can lead to tangible economic benefits for hotels.

Concluding remarks, implications and future developments

Main results and contributions

This study proposes an investigation of the causal relationship running from topics in online reviews to customer satisfaction, passing through sentiments and emotions, expressed by tourists and referred to their stay in Sardinian hotels. The main idea is to use natural language processing to understand whether there are differences between coastal and inland hotels in topics used by online reviewers and whether the same topics might show different impacts between locations. To retrieve and process data from the web, we apply an adapted version of the new method recently proposed by[6]. The main results of this study underscore the significant role of online customer reviews in explaining customer satisfaction, particularly in the context of hotel ratings in Sardinia (RQ1). We have identified specific topics in online reviews that positively or negatively influence these ratings (RQ2), with notable differences in the impact of these topics between coastal and inland hotels (RQ3). Our findings indicate that different elements of a review, including topics, sentiments, and emotions, contribute variably to the customer's final satisfaction score. This suggests that customer satisfaction is not solely based on the physical attributes of the hotel, but also on the emotional experiences reflected in the reviews. Furthermore, the application of the TOBIAS method allowed us to quantify the weight of these elements, providing a nuanced understanding of the factors driving satisfaction in different geographic locations. This distinction between coastal and inland hotels is crucial for tailoring marketing and management strategies. According to the two-factors theory, we find that the final tourist rating assessment depends on a combination of both satisfaction (Positiveness) and dissatisfaction (Negativeness) which are contingent upon the sentiment and emotions evoked by the tourism experience.

With these results, the present investigation presents noteworthy and diverse contributions to the tourism literature. A primary innovation pertains to the utilization of a recent developed methodology hitherto unexplored in tourism research (such as the combination of topic modeling, sentiment analysis and PLS-PM), aimed at examining the relationship between online reviews and ratings in the hospitality sector. This study represents a pioneering endeavor in estimating the impact of predominant topics emerging evident in reviews on the tourist satisfaction expressed by the rating, and in establishing weights and signs of each of the topics. Additionally, a significant contribution lies in the comparison between coastal and inland hotels. Particularly within the context of Sardinia, this comparative analysis assumes importance in fostering growth and development in areas. In a broader sense, this work improves the literature on customer satisfaction by introducing novel analytical tools and perspectives.

Moreover, our results highlight how some issues, like room services, can be directly addressed by hotel managers, while others, like destination parking, require public intervention. Thus, the analysis points to a relevant role of policymakers and two levels of possible intervention. Policymakers are urged to enhance infrastructure and services for overall destination management, while also focusing on specific amenities crucial for accommodation facilities. In this respect, stakeholder preferences gleaned from online reviews may represent an extraordinary resource enabling policymakers to craft effective long-term strategies for tourism development.

Managerial and policy implications

Such in-depth insights pave the way for intriguing possibilities: for the current study, it enables the drawing of significant managerial and administrative policy recommendations, while, for future research, it offers a robust framework for analyzing a broad spectrum of economic activities.

In other words, our results prove that TOBIAS is an appropriate and useful tool for management analysis aimed at developing and improving tourism supply. The strength of this method is that it can be applied in different contexts, for both more specific facilities and the destination as a whole. By identifying the specific topics that impact tourist satisfaction, it becomes possible for managers and policymakers to tailor specific intervention policies. In the specific case of inland hotels, this analysis highlights a need for improvement in what concerns the quality of the room and the connected services (such as shower, Wi-Fi, towels-hairdryer). At the same time, it reveals how emotions and sentiments strongly contribute to determining the negativity or positivity of the quality assessment. These results concern fundamental aspects of hospitality services to which much attention should be paid. Moreover, the differences recorded between inland and coastal hotels highlight how tourist needs change with respect to the location of the hotel and, consequently, to the kind of vacation they are realizing. At the same time, the results highlight how some of the topics mentioned in the reviews can be directly improved by the hotel manager, such as service related to the room and the breakfast, while others can be improved only by the policymaker and the public intervention, as for instance the issues related to the parking at destination. As regards policymakers, our analysis highlights two levels of possible intervention. In general, the integrated development of inland tourism requires a greater effort towards improving infrastructure and services for the management of the tourism destination as a whole (such as transport). Furthermore, policymakers should pay close attention to the development of some specific services to support accommodation facilities such as the internet, car parks or insect disinfestation, and so on. In this context, policymakers can craft decisions and develop long-term strategies informed by a knowledge base that reflects stakeholder preferences, as articulated in online reviews. In the context of this research, the TOBIAS method emerges as a powerful tool for analyzing management strategies in tourism, applicable to various contexts from specific facilities to entire destinations. By identifying the role of specific topics, it enables tailored interventions by pinpointing factors affecting tourist satisfaction. In this respect, variances between inland and coastal hotels underscore the need for location-specific approaches.

Limitations and future research

Despite its contributions, this study has two main limitations regarding the sampling area and the electronic platforms used. These have certainly limited the range of topics that emerged from the reviews and reduced the extendibility of our results to other contexts. Additionally, TOBIAS focuses on sentiment and topic modeling to explain customer satisfaction, but it does not account for all possible factors that might influence satisfaction, such as personal preferences, cultural differences, or specific situational contexts that are not captured in textual reviews. To mitigate the impact of confounding factors, we employed several strategies. First, we used advanced natural language processing (NLP) techniques to extract and analyze a comprehensive set of emotional features and topics from the reviews, ensuring a robust representation of sentiments and themes. Second, we applied Partial Least Square Path Modeling (PLS-PM) to model the relationships between topics, sentiments, and overall satisfaction, which allowed us to control for the influence of various factors simultaneously. Third, we distinguished between coastal and inland hotels, recognizing the potential for different customer expectations and experiences based on location, which helped isolate the effects of specific contextual variables. These methodological choices aimed to enhance the reliability of our findings and reduce the impact of potential biases inherent in the data. However, the TOBIAS model offers a robust framework for analyzing a broad spectrum of economic activities. It can be easily replicated in other geographical areas and for specific services, by exploiting other web platforms. Therefore, the limitations of this investigation could potentially represent an opportunity for new studies in our future research pipeline.

Data availability

The datasets analyzed during the current study are not publicly available due to privacy reasons, but are available from the corresponding author on reasonable request.

Received: 12 June 2024; Accepted: 30 September 2024
Published online: 09 October 2024

References

1. Disegna, M. & Osti, L. Tourists' expenditure behaviour: The influence of satisfaction and the dependence of spending categories. *Tour. Econ.* **22**, 5–30 (2016).
2. Sukiman, M. F., Omar, S. I., Muhibudin, M., Yussof, I. & Mohamed, B. Tourist satisfaction as the key to destination survival in Pahang. *Procedia Soc. Behav. Sci.* **91**, 78–87 (2013).
3. Zhou, K. & Yao, Z. Analysis of customer satisfaction in tourism services based on the Kano model. *Systems* **11**, 345 (2023).
4. Bigne, E., Ruiz, C., Perez-Cabañero, C. & Cuenca, A. Are customer star ratings and sentiments aligned? a deep learning study of the customer service experience in tourism destinations. *Serv. Bus.* **17**, 281–314 (2023).
5. Yin, R. K. *Case study research: Design and methods* Vol. 5 (Sage Publications, 2009).
6. Ortu, M., Frigau, L. & Contu, G. Topic based quality indexes assessment through sentiment. *Comput. Stat.* **39**, 289–311 (2024).
7. Chen, C.-H., Nguyen, B., Klaus, P. P. & Wu, M.-S. Exploring electronic word-of-mouth (eWOM) in the consumer purchase decision-making process: The case of online holidays-evidence from United Kingdom (UK) consumers. *J. Travel Tour. Mark.* **32**, 953–970 (2015).
8. Dhar, S. & Bose, I. Walking on air or hopping mad? understanding the impact of emotions, sentiments and reactions on ratings in online customer reviews of mobile apps. *Decis. Support Syst.* **162**, 113769 (2022).
9. Chen, C.-M., Chen, S. H. & Lee, H. T. The destination competitiveness of Kinmen's tourism industry: Exploring the interrelationships between tourist perceptions, service performance, customer satisfaction and sustainable tourism. *J. Sustain. Tour.* **19**, 247–264 (2011).
10. Cuomo, M. T. *et al.* Digital transformation and tourist experience co-design: Big social data for planning cultural tourism. *Technol. Forecast. Soc. Chang.* **162**, 120345 (2021).
11. Cheung, C. M., Lee, M. K. & Rabjohn, N. The impact of electronic word-of-mouth: The adoption of online opinions in online customer communities. *Internet Res.* **18**, 229–247 (2008).
12. Wen, I. Factors affecting the online travel buying decision: A review. *Int. J. Contemp. Hosp. Manag.* **21**, 752–765 (2009).
13. Xie, K. L., Zhang, Z. & Zhang, Z. The business value of online consumer reviews and management response to hotel performance. *Int. J. Hosp. Manag.* **43**, 1–12 (2014).
14. Calheiros, A. C., Moro, S. & Rita, P. Sentiment classification of consumer-generated online reviews using topic modeling. *J. Hosp. Mark. Manag.* **26**, 675–693 (2017).
15. Geetha, M., Singha, P. & Sinha, S. Relationship between customer sentiment and online customer ratings for hotels-an empirical analysis. *Tour. Manag.* **61**, 43–54 (2017).
16. Kim, J. & Im, J. Proposing a missing data method for hospitality research on online customer reviews: An application of imputation approach. *Int. J. Contemp. Hosp. Manag.* **30**, 3250–3267 (2018).
17. Ameur, A., Hamdi, S. & Ben Yahia, S. Sentiment analysis for hotel reviews: A systematic literature review. *ACM Comput. Surv.* **56**, 1–38 (2023).
18. Nguyen, V.-H. & Ho, T. Analyzing customer experience in hotel services using topic modeling. *J. Inf. Process. Syst.* **17**, 586–598 (2021).
19. Al-Natour, S. & Turetken, O. A comparative assessment of sentiment analysis and star ratings for consumer reviews. *Int. J. Inf. Manag.* **54**, 102132 (2020).
20. Chang, J.-R., Chen, M.-Y., Chen, L.-S. & Tseng, S.-C. Why customers don't revisit in tourism and hospitality industry?. *Ieee Access* **7**, 146588–146606 (2019).
21. Shah, S. & Bhatt, D. A study on tourist satisfaction with the services provided by the hotel industry of Naintal. *Int. J. Tour. Hotel Bus. Manag.* **2**, 202–217 (2020).
22. Chen, X. & Xie, H. A structural topic modeling-based bibliometric study of sentiment analysis literature. *Cogn. Comput.* **12**, 1097–1129 (2020).
23. Ali, T., Omar, B. & Soulaimane, K. Analyzing tourism reviews using an LDA topic-based sentiment analysis approach. *MethodsX* **9**, 101894 (2022).
24. Suryadi, D. & Imran, J. A. A topic modeling and sentiment analysis approach for benchmarking of hotels based on online reviews. *Ind. Eng. Manag. Syst.* **21**, 646–657 (2022).
25. Ounacer, S., Mhamdi, D., Ardchir, S., Daif, A. & Azzouazi, M. Customer sentiment analysis in hotel reviews through natural language processing techniques. *Int. J. Adv. Comput. Sci. Appl.* **14**, 569–579 (2023).
26. Akhtar, N., Zubair, N., Kumar, A. & Ahmad, T. Aspect based sentiment oriented summarization of hotel reviews. *Procedia Comput. Sci.* **115**, 563–571 (2017).
27. Zhu, L., Lin, Y. & Cheng, M. Sentiment and guest satisfaction with peer-to-peer accommodation: when are online ratings more trustworthy?. *Int. J. Hosp. Manag.* **86**, 102369 (2020).
28. Kim, Y.-J. & Kim, H.-S. The impact of hotel customer experience on customer satisfaction through online reviews. *Sustainability* **14**, 848 (2022).
29. Oliver, R. L. A cognitive model of the antecedents and consequences of satisfaction decisions. *J. Mark. Res.* **17**, 460–469 (1980).
30. Lazarus, R. S. *Emotion and adaptation* (Oxford University Press on Demand, 1991).
31. Herzberg, F., Mausner, B. & Snyderman, B. B. *The motivation to work* (Wiley, 1959).
32. Zhang, Z., Li, H. & Law, R. Differences and similarities in perceptions of hotel experience: The role of national cultures. *J. Travel Tour. Mark.* **32**, S2–S14 (2015).
33. Assaker, G., Hallak, R., O'Connor, P. & Esposit Vinzi, V. Partial least squares path modelling (PLSPM): A new direction for research in tourism and hospitality. *J. Travel Tour. Res.* **13**, 158 (2013).

34. Rahmadan, M. C., Hidayanto, A. N., Ekasari, D. S., Purwandari, B. *et al.* Sentiment analysis and topic modelling using the lda method related to the flood disaster in Jakarta on twitter. In *2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, 126–130 (IEEE, 2020).
35. Liu, Y. *et al.* Roberta: A robustly optimized bert pretraining approach. Preprint at arXiv:1907.11692 (2019).
36. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. Preprint at arXiv:1810.04805 (2018).
37. Grootendorst, M. Bertopic: Neural topic modeling with a class-based tf-idf procedure. Preprint at arXiv:2203.05794 (2022).
38. Dijkstra, T. K. Latent variables and indices: Herman Wold's basic design and partial least squares. In *Handbook of partial least squares: Concepts, methods and applications* 23–46 (Springer, 2009).
39. Benitez, J., Henseler, J., Castillo, A. & Schuberth, F. How to perform and report an impactful analysis using partial least squares: Guidelines for confirmatory and explanatory is research. *Inf. Manag.* **57**, 103168 (2020).
40. Sotiriadis, M. D. Management and operational issues of animation services in resort and all-inclusive hotels: Evidence from Greece. *Mediterr. J. Soc. Sci.* **5**, 692–698 (2014).
41. Keča, K. & Ivančić, I. The influence of the offer of animation programs on customer satisfaction and loyalty. In *DIEM: Dubrovnik International Economic Meeting*, vol. 7, 80–90 (Sveučilište u Dubrovniku, 2022).
42. Smirnova, E. Clean room, uncomfortable bed: A corpus analysis of evaluation devices in hotel reviews. *Pragmat. Soc.* **13**, 193–223 (2022).
43. Albayrak, T. & Caber, M. Prioritisation of the hotel attributes according to their influence on satisfaction: A comparison of two techniques. *Tour. Manag.* **46**, 43–50 (2015).
44. Liu, R., Zhang, Y., Luo, C., Tan, S. & Gong, Y. Review content type and hotel review helpfulness: Direct and moderating effects. *Inf. Technol. Manag.* https://doi.org/10.1007/s10799-023-00392-0 (2023).
45. Wang, L., Wang, X.-K., Peng, J.-J. & Wang, J.-Q. The differences in hotel selection among various types of travellers: A comparative analysis with a useful bounded rationality behavioural decision support model. *Tour. Manag.* **76**, 103961 (2020).
46. Soler, I. P., Gemar, G. & Guzman-Parra, V. F. A comparison of destinations' impacts on hotel rates. *Int. J. Hosp. Manag.* **77**, 226–237 (2019).
47. Assaf, A. G., Josiassen, A., Woo, L., Agbola, F. W. & Tsionas, M. Destination characteristics that drive hotel performance: A state-of-the-art global analysis. *Tour. Manag.* **60**, 270–279 (2017).
48. Phillips, P., Barnes, S., Zigan, K. & Schegg, R. Understanding the impact of online reviews on hotel performance: An empirical analysis. *J. Travel Res.* **56**, 235–249 (2017).
49. Anderson, C. *The impact of social media on lodging performance* (2012).
50. Kim, W. G. & Park, S. A. Social media review rating versus traditional customer satisfaction: which one has more incremental predictive power in explaining hotel performance?. *Int. J. Contemp. Hosp. Manag.* **29**, 784–802 (2017).
51. Ali, F. & Amin, M. The influence of physical environment on emotions, customer satisfaction and behavioural intentions in Chinese resort hotel industry. *J. Glob. Bus. Adv.* **7**, 249–266 (2014).

Author contributions

M.O. and G.C. and C.M contributed to the conceptualization. M.O. and G.C. contributed to the methodology. C.M and C.D. supervised the manuscript. All authors wrote the main manuscript text. All authors reviewed the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.O.

Reprints and permissions information is available at www.nature.com/reprints.