

Article

Not peer-reviewed version

Experience Economy Perspective on Recreational Fishing Tourism Travelers' Reviews: A Data Science Approach

[Georgia Yfantidou](#)*, Sofia Yfantidou, Dimitra Karanatsiou, [Angelos Lontakis](#), Mairi Maniopoulou, Panagiota Balaska, [Alkistis Papaioannou](#), Vassiliki Vassilopoulou

Posted Date: 1 March 2024

doi: 10.20944/preprints202403.0041.v1

Keywords: fishing tourism; sustainable tourism; recreation; experience economy; TripAdvisor; user profiling; Natural Language Processing



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Experience Economy Perspective on Recreational Fishing Tourism Travelers' Reviews: A Data Science Approach

Georgia Yfantidou ^{1,*}, Sofia Yfantidou ², Dimitra Karanatsiou ², Angelos Liontakis ³, Mairi Maniopoulos ⁴, Panagiota Balaska ⁵, Alkistis Papaioannou ⁶ and Vassiliki Vassilopoulos ⁴

¹ Department of Physical Education and Sport Science, Democritus University of Thrace

² School of Informatics, Aristotle University of Thessaloniki

³ Department of Agribusiness and Supply Chain Management, Agricultural University of Athens

⁴ Hellenic Center for Marine Research, Institute of Marine Biological Resources and Inland Waters

⁵ Department of Physical Education and Sport Science, Aristotle University of Thessaloniki

⁶ School of Social Sciences, Hellenic Open University

* Correspondence: gifantid@phyed.duth.gr; Tel.: +30-25310-39709

Abstract: This study explores the Experience Economy dimensions within fishing tourism, adopting a data science perspective by employing web crawling for extensive data gathering on reviewer profiles from TripAdvisor. Utilizing Natural Language Processing (NLP) techniques, it scrutinizes the relationship between Experience Economy dimensions and user profiling aspects. Findings reveal "Entertainment" as the predominant dimension in tourist reviews, followed by "Aesthetic," "Educational," and "Escapist." Particularly notable are the frequent co-occurrences of "Entertainment"- "Aesthetic" and "Educational"- "Entertainment" pairs. The practical implications underscore the potential socioeconomic benefits of fishing tourism for local communities and fishermen, emphasizing the necessity for governmental support in terms of infrastructure, leadership, legislation, and financial backing to foster sustainable development in this sector. Notably, this research stands as a pioneering effort in its focus on Experience Economy dimensions and user profiling within the context of fishing tourism, drawing data from business pages and user profiles.

Keywords: fishing tourism; sustainable tourism; recreation; experience economy; TripAdvisor; user profiling; Natural Language Processing

1. Introduction

Fishing tourism is a developing form of recreational tourism that promotes fisheries and aquaculture, offering travelers an opportunity to enrich their activities. Fishing tourism can be defined as "a set of activities carried out by professionals in order to differentiate their incomes, promote and valorize their profession and socio-cultural heritage, and enhance sustainable use of marine ecosystems, by means of boarding non-crew individuals on fishing vessels" [1]. According to a 2020 study by the Centre for the Promotion of Imports from Developing Countries of the Dutch Ministry of Foreign Affairs, "The European Potential for the Development of Sports Tourism", sports tourism was the fastest growing form of tourism before the pandemic. After the pandemic we can expect a lot of opportunities, especially in certain sports that contribute to sustainable development. This study showed that the sports tourism market can be divided into nine smaller markets. Figure 1 illustrates a model describing the smaller focused sports tourism niche markets that contribute to sustainable development. The model refers to those groups of tourists who have a strong commitment to their sport of choice and common characteristics (high education and income, active in social networks, belong to smaller age groups, etc.) [2].

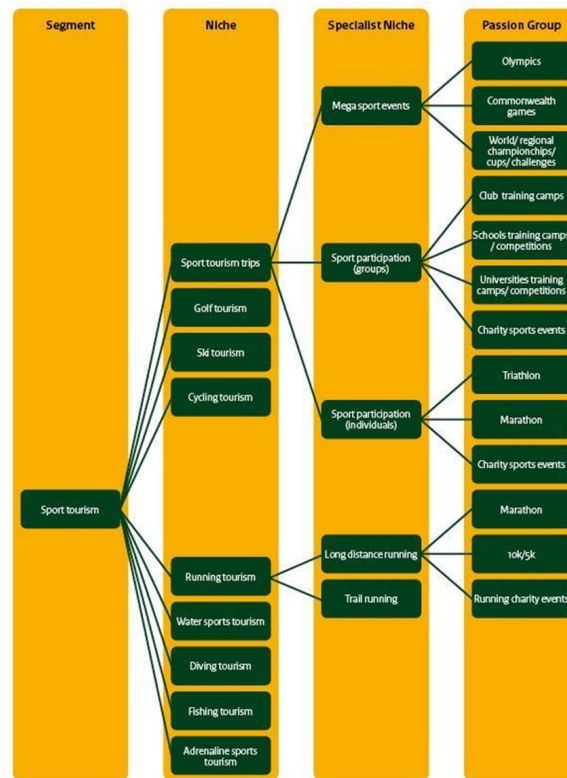


Figure 1. Niche markets of sport tourism, which include fishing tourism based on the research of CBI (Figure courtesy of the Centre for the Promotion of Imports from developing countries [2]).

Travelers often seek experiences in different tourism types, for instance, food and wine tourism, sports tourism, or luxury tourism, based on specific purposes. Desirable experiences can positively affect emotions, remain in customers' memory, and influence their consequent behaviors. Hence, understanding customer experience in the tourism sector is crucial for building better products and services. As such, we utilize the Experience Economy perspective [3] to explore different experience realms. The Experience Economy presents four realms, the so-called 4Es, of experiential value for a business. The 4Es add Educational, Esthetic, Escapist, and Entertainment experiences to a business offering.

In line with the above, the main task of this project is to classify tourism dimensions and build tourist profiles based on text reviews for fishing tourism businesses on Tripadvisor. The project is conducted in cooperation with the Hellenic Centre for Marine Research (HCMR) and the Department of Physical Education & Sport Science of the Democritus University of Thrace (DUTH). Its results will help to better understand this new form of eco-tourism and further bolster fishermen's livelihoods. Our first goal is to classify tourists' reviews on TripAdvisor according to the Experience Economy perspective to increase our understanding of the tourist experience regarding fishing tourism and recreation; while exploring the topics that emerge within these dimensions. At the same time, we focus on detecting tourist profiles based on their reviews, not only about fishing tourism businesses but also other businesses they might have reviewed during their travels. This sort of analysis can provide the HCMR and local island businesses with additional knowledge of their customer base and might enable them to engage in destination development through niche tourism. Niche tourism refers to "how a specific tourism product can be tailored to meet the needs of a particular audience/market segment" [4]. Specifically, our second goal is to examine the feasibility of building reviews-based predictors of various user characteristics, including, but not limited to, gender, age group, marital status, and

interests. The latter category is particularly interesting because it allows us to explore secondary tourist interests to create more holistic experiences.

2. Related Work

Applying Natural Language Processing (NLP) methods to the tourism domain is no new undertaking. Online platforms, such as Yelp and TripAdvisor, offer millions of publicly available reviews of businesses in the tourism domain, making them an attractive data source for research projects. However, to the best of our knowledge, this is the first study that aims to explore the Experience Economy dimensions under in relation to user profiling aspects on fishing tourism.

2.1. Fishing Tourism

Marine tourism is the set of recreational activities and experiences that take place in the marine and coastal areas of a country to provide entertainment to tourists. Fishing tourism is also part of this category, which means that it can offer services to visitors interested in alternative activities and experience tourism and is gaining ground in the preferences of tourists who love fishing and want to experience it. Fishing tourism is defined by the promotion and exploitation of fishing and aquaculture, activities with a rich traditional character, in terms of employment and the means used, but also in terms of the aquatic environment and aquatic life. In particular, fishing tourism is defined as:

- the performance of the daily fishing process, accompanied by an explanation of the process to passengers
- encouraging the active, safe participation of visitors in the whole process of fishing and with the opportunity to be engaged at marine sport activities
- informing tourists about the fishing activity and fishing tradition
- visits to beaches, underwater caves, and boat trips
- possibility of diving for fishing and observation of marine flora and fauna
- contact with local flavors and traditional cooking of the catch
- on-site tasting and sale of traditional fishing products
- overnight accommodation and catering services in fishermen's houses or other "fisherman's style" establishments

Fishing tourism is developed in all aquatic destinations of the country, including seas, lakes, rivers, and lagoons, where fishing, farming, and breeding of aquatic organisms can be practiced. Furthermore, this form of tourism creates new infrastructure and jobs and is characterized by three forms: a) active, b) passive, and c) shore fishing. In the active mode, the tourist actively participates in the fishing activity through a privately owned or chartered boat. In passive fishing, the tourist boards a professional fishing boat and watches the fishing activity as a spectator, thus coming into contact with the natural environment. Finally, shore fishing is a very popular recreational activity that involves long periods of inactivity that allow for rest and relaxation and also offers periods of action in case the fish is hooked.

The experience of trying to catch the fish, the enjoyment of escaping from work and everyday life, is generally more important than the actual event of catching a large quantity of fish. In many areas, shore fishing has been organized to take the form of a tourist product. This tourist product, i.e., "shore fishing", involves firstly the sale of fishing licenses for a limited period of time, the renting of rooms in tourist accommodation in the area by tourists who come exclusively for the purpose of fishing and the provision of guiding services for the fishing area. In order not to exceed the bearing capacity of the area, the number of amateur fishermen is strictly controlled, and the amateur fishermen follow the tactic of catch and release. The terms and conditions for carrying out fishing tourism must be exercised by professional fishermen and owners of professional vessels who wish to carry out fishing tourism alongside their professional fishing activities. Fishing vessels carrying out fishing tourism are required to:

- Have an overall length of up to 15 meters.

- They must be equipped with a professional fishing license for fishing with gear other than bottom trawls with nets and boat-drawn gillnets.
- They must meet the requirements of professional tourist vessels under the relevant laws.
- Carry up to 12 passengers.
- Be equipped with a certificate of seaworthiness stating the number of passengers they can carry, the extent of the voyages, and the relevant “Orders - Instructions” without requiring the issue of special or other certificates.
- There shall be a special waiting area for all passengers to be safely accommodated during fishing operations without obstructing them.
- Comply with the rules laid down by the legislation in force at the time concerning the safety of navigation, manning, hygiene, and the suitability of the fishing vessel for the embarkation of passengers.

When conducting fishing tourism, professional fishermen or sponge fishermen shall demonstrate fishing or sponge fishing techniques in accordance with the national and fisheries legislation in force, using the fishing methods and gear specified in the vessel’s professional license, except for bottom trawling with gillnets and boat-towed gillnets. Furthermore, fishing gear shall be so arranged on board the vessel that it does not impede the free and safe movement of the passengers and any activity on board. Furthermore, tourists may fish only with fishing lines, trolling lines, and probes, which may be handled manually and not mechanically, and may participate, under the responsibility of the master of the vessel and during fishing activities, only in operations that do not endanger their safety. Fishing tourism, for fans and non-fans alike, is an unprecedented and exciting activity that goes beyond the usual and introduces you to a different form of tourism and presentation of the beauty of a destination accompanied with recreation.

2.2. *The Experience Realm*

Holbrook and Hirschman [5] and Pine et al. [6] are the first authors of business studies to interpret in their book on business management the categories of experience “4E”. They present the nature of the experiences in terms of economic activities. Pine & Gilmore’s groundbreaking work (1998) illustrates the four ways in which customers (tourists) can become involved or engaged in tourism experiences. Coupling of the dimension “tourist participation” with the dimension “environmental relationship” defines the four “realms” of an experience:

1. **Entertainment:** Usually, this experience is passively gained where the viewer is not directly involved in the “performance” of the entertainment. e.g., participation in the theatre, cinema, concerts, parades, nightclubs, etc. carnivals and carnivals, and folklore festivals as spectators.
2. **Education:** This type is the result of active participation and absorption of the material element that a person has been exposed to. The presentation by speech at a conference of thematic modules that are simultaneously a professional e.g., a doctor, can be considered an experience of this module classification.
3. **Aesthetics:** The category of this type of experience is based on both exciting and passive enjoyment. A classic example of this type of experience is the understanding and inner search for the stimuli evoked by a series of specific themes of artistic works that are exhibited in a gallery or in an exhibition of unique exhibits in a museum.
4. **Escape:** This type exists when you are immersed in an activity that is actively engaged in by stakeholders who are transported into a new state of experience, e.g., role-playing to enhance relationship building in a working group that as partners take on the role of role models playing the role of experts to solve a crisis problem, e.g., due to an epidemic.

The shift towards prioritizing customer experiences and emotions stems from a changing consumer landscape, where passive engagement is no longer sufficient. Instead, customers actively seek meaningful interactions and value-added propositions. This evolution necessitates a departure from

traditional sales approaches towards fostering engagement and co-creation. Concepts such as imagination, participation, and co-creation serve as vehicles for delivering this new paradigm of value, where customers play an integral role in shaping the products and services they consume. The theoretical framing of this experiential development in marketing came through the [6] proposal of 4E theory whereby experiences are classified between two axes and the poles of these axes as shown in Figure 2 [7]:

- Active (act) and Passive (accept) participation, respectively, in the first axis and
- Basic adaptation/absorption and Total immersion, respectively, in the second axis

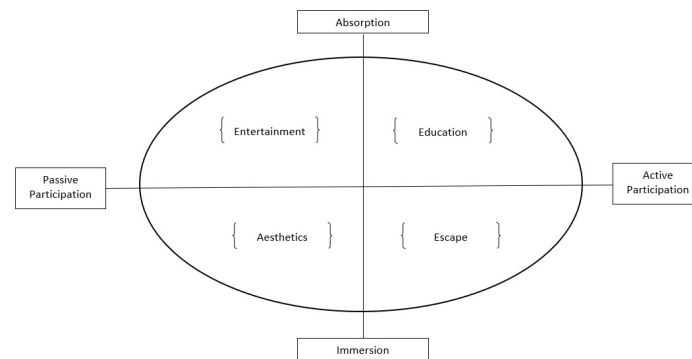


Figure 2. The experience realm (Figure courtesy of Pikkemaat and Weiermair [7]).

2.3. Natural Language Processing and the Tourism Domain

Topic detection on travelers' reviews has been one of the main focuses of researchers in the tourism field [8,9]. Afzaal et al. [10] developed a multi-aspect classification approach, identifying aspects such as food, price, location, service, and ambiance on reviews from multiple online platforms. The same group [11] developed an alternative approach, including also the sentiment of the comment in the classification process. Sentiment Analysis methods have also been applied in the field of online tourist reviews. Yu et al. [12] focused on foreign language sentiment analysis focusing on Japanese tourists' reviews. Marrese-Taylor et al. [13] introduced new domain-specific features for sentiment extraction, while Kirilenko et al. [14] provided a comparison between different Machine Learning algorithms for the task of sentiment analysis in the tourism domain. However, none of the above approaches have applied topic modeling, and sentiment analysis approaches to examine the concept of user experience and niche tourism, as well as the tourists' perception of those concepts. More user-oriented research has focused on predicting reviews' usefulness [15,16], identifying suitable attractions' recommendations for users [17,18], and extracting certain user profiles [19]. While these works study individual user behavior, they do not examine the correlation between user profiles and tourist experiences to better understand different market segments. To the best of our knowledge, our work is the first to study user experiences, behaviors, and profiles in the domain of niche tourism.

2.4. User Profiling in Tourism-Related Platforms

Recommender systems are mainly based on user choices and profiles to make accurate recommendations. These profiles are created in different ways regarding the field of application, for example, based on market basket [20] or user-generated-content in the form of images, videos [21] or reviews [22]. In the tourism industry, recommendation engines on relevant platforms build individual tourist profiles to (a) suggest hotels, restaurants, attractions or routes based on the shared ratings, reviews, photos, videos, or likes and (b) provide businesses with insights about their customers segments. Recent studies focus on tourist reviews to build user profiles. Kavitha et al. [23] exploit TripAdvisor reviews and social media profiles metadata to build a destination recommendation engine based on users' previously visited locations and matching user experiences about a destination. Moreover, Leal et al. [17] designed an algorithm for personalized destinations recommendation based on Expedia

reviews by applying content-based filtering to topic-modeled tourists and locations. Except for the destination recommendation itself, researchers also focus on tourism-related services, i.e., restaurants, activities, etc. The approach of Missaoui et al. [24] utilizes users' Yelp reviews to recommend the most relevant services by taking into account the opinions that this user has explicitly expressed through her/his previous reviews concerning other similar services following a language modeling approach. In addition, patterns on travelers' preferences extracted from reviews have also been identified by Fazzolari and Petrocchi [25], who proposed methods that automatically analyze and summarize the reviews' features. However, none of the above methods inferred specific profiles from users' review history and platform metadata. To the extent we are aware of, our work is the first to build user profiles with targeted attributes-categories (i.e., gender, age, marital/family status, interest in activities, etc.) that successfully meet the needs of niche tourism businesses.

3. Materials and Methods

This sections discusses the process of acquiring and annotating the ground truth data prior to linguistic analysis.

3.1. Pipeline Overview

The raw, unlabeled data for the project are crawled from (1) fishing tourism businesses' pages on TripAdvisor and (2) the respective user profiles on TripAdvisor. Initially, the HCMR researchers provide the list of fishing businesses' pages on TripAdvisor. After crawling the business' pages to get user reviews, we need to label our data with the 4Es labels to train the respective classifiers. Along with the HCMR researchers, we manually label a subset of the dataset based on a specific methodology that includes word frequencies of specific, pre-defined keywords. Then, for each collected review, we crawl the reviewer's profile data. These data contain information about previously visited locations and reviews as well as some demographic data (if available), such as the user's gender, age, interests, and permanent (home) location. Note that all data are pseudonymized upon collection. In addition, TripAdvisor assigns specific badges to users as a recognition for their contribution to the platform. These badges refer mainly to the platform use, for instance, if the user attracts readership attention or writes multiple reviews. However, some of these badges reflect the interests and behavior of the user while traveling, in particular expertise in restaurant reviews, in specific types of visited accommodation (luxury, resort, etc.), in attractions, in photography, and many more. All the aforementioned data (explicit profile) will be aggregated with specific characteristics mined from the reviews (implicit profile) to create holistic and informative tourist profiles.

3.2. Dataset Collection

TripAdvisor does not grant access to its content Application Programming Interface (API) for academic research purposes. Thus, to collect data for this project, we resort to web crawling as a means of mass data collection. A web crawler "is an Internet bot that systematically browses the World Wide Web, typically for the purpose of Web indexing"¹. In other words, it imitates the behavior of a user by visiting a set of pages and extracting data from them. Our web crawler operates per the *robots.txt* file of TripAdvisor; namely, it only requests pages that are permitted according to the file's directives.

For the scope of this project, we crawl the TripAdvisor profiles of 30 fishing businesses from various Greek islands, containing 1062 English reviews from 1026 unique users. For each review, we store the reviewer's profile link and the date, rating, title, and full text of the review. To extract these data, we utilize the Beautiful Soup², a Python library for pulling data out of HyperText Markup Language (HTML) and Extensible Markup Language (XML) files, along with regular expressions.

¹ https://en.wikipedia.org/wiki/Web_crawler

² <https://www.crummy.com/software/BeautifulSoup/>

Then, for each of the 1018 reviewers, we crawl their personal profiles, namely reviews, demographics, and badges. Precisely, we extract 11862 user reviews for 9071 different businesses in 2497 locations worldwide. At the same time, for each of the 1018 users, we collect a sum of 11562 badges of 27 unique types, such as “Hotel Expert” or “New Photographer”. Finally, for 999 users whose demographics are fully or partially available, we collect data regarding gender (158 users), age (103 users), location (597 users), membership year (997 users), number of cities visited (706 users), number of contributions to TripAdvisor (687 users), and related tags (5 users). To enable the dynamic crawling of user profiles, we use Selenium, a portable framework for testing web applications³, since the functionality of BeautifulSoup is not sufficient for this task.

Summing up, we build a new dataset of user reviews and related user data focused on niche tourism and specifically fishing tourism by crawling (1) fishing tourism businesses’ pages on TripAdvisor and (2) the respective user profiles on TripAdvisor.

3.3. Dataset Annotation

Ground truth is required to apply supervised learning techniques to classify tourists’ reviews on TripAdvisor according to the Experience Economy perspective. To acquire such ground truth, we label a subset (200 comments) of the first dataset, i.e., reviews for fishing tourism businesses, where each comment is annotated by two to three researchers. Each comment is labeled with one or more of the 4Es dimensions, making our task a multi-label classification problem. The labeling decision is based on the domain expertise of the researchers who have created a list of key phrases for each dimension of the Experience Economy. For example, the Education dimension may contain concepts such as fishing techniques (e.g., casting, UK-style fishing, kayak fishing) or responsible fishing, while the Entertainment dimension may contain concepts such as swimming, eating, or cooking. Similar concepts have been identified for the remaining two dimensions. For the second dataset, i.e., user profiles, we utilize the scraped demographics data, such as gender or age, as labels for supervised learning tasks.

4. Results

In this section we delineate our findings on extracting sentiments, emotions and relevant insights from the reviews’ linguistic cues. We also explore user profiling aspects in fishing tourism and how the 4Es translate to this niche tourism product.

4.1. Extracting Sentiment, Emotion, and Descriptive Insights of Tourists and Businesses from Linguistic Cues

To gain deep knowledge and insights into customers’ collective word of mouth for fishing businesses and the tourists’ general interests, we performed natural language processing tasks on the extracted reviews. In particular, we first preprocess the text of the reviews to find the most occurring words and phrases and relevant linguistic insights. Moreover, we carry out sentiment and emotion analysis following mainly an unsupervised approach using relevant lexicons and relevant expressions. Finally, we applied text clustering and topic extraction to the reviews to elicit specific interests of tourists and commonly emerging user experiences.

4.1.1. Linguistics Insights

Text cleaning. The initial step is to clean the text of the reviews in order to extract noisy info and maintain a clear dataset that will be used for future tasks. We first converted the words in our corpus to lowercase. Then we filtered out external URLs, Unicode characters and emojis (symbols and pictographs), digits, and punctuation. After this step, all *stopwords*, the set of most commonly used in the language in general, were taken out of the corpus.

³ <https://www.selenium.dev/>

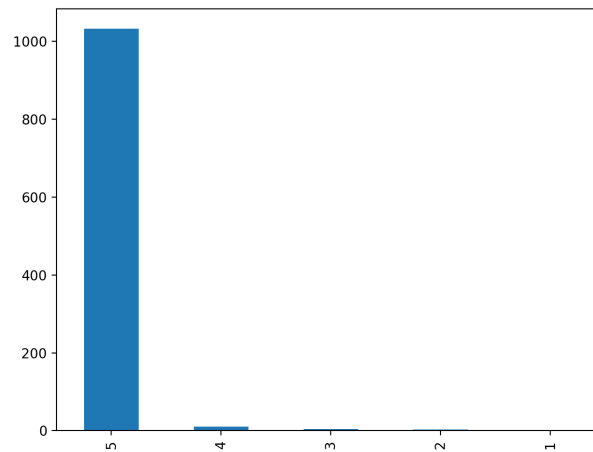


Figure 4. User ratings that appear in fishing businesses reviews (Source: Authors own creation).



Figure 5. Words and concepts that appear most in users' reviews (Source: Authors own creation).

An interesting finding is that the rating distribution for overall user reviews (Table 1) significantly differs from the one identified previously in the fishing businesses' reviews (Figure 4). Specifically, while only 7% of the overall user reviews have a 5-star rating, this percentage rises to almost 100% when it comes to fishing tourism.

Table 1. General user ratings distribution and insights (Source: Authors own creation).

Rating	Counts-percentage	avg #words
1	6087 (51.3%)	82.3
2	15 (0.1%)	59.1
3	46 (0.3%)	118.4
4	4795 (40.4%)	95.0
5	919 (7.0%)	81.8

Additionally, we have explored the TripAdvisor badges assigned by TripAdvisor to each user to highlight one's contribution to the community. Specifically, TripAdvisor assigns badges for the following categories:

- **Reviewer Badges:** These badges are graded starting from the "New Reviewer" (1 review) to the "Top Contributor" (more than 50 reviews). Figure 6a shows the distribution of tourists who reviewed fishing tourism business based on their "Reviewer Badges". Interestingly, approximately half the users belong to the "New Reviewer" category, meaning that they only joined the TripAdvisor platform to positively review the respective fishing tourism business.

- **Expertise Badges:** These badges showcase the unique knowledge of the users. For example, if a user publishes multiple reviews in a single category – hotels, restaurants, or attractions - they will be assigned the respective “Expertise Badge”. Figure 6b shows the distribution of tourists who reviewed the fishing tourism business based on their “Expertise Badges”. We notice that the users who review fishing tourism businesses also tend to review hotels and attractions, as well as luxury and boutique hotels, and B&B and Inns at a smaller scale.
- **Passport Badge:** This badge recognizes users for being world travelers. Once they have added reviews for places in at least two destinations, they start collecting such graded badges. Figure 7 shows that most users who have reviewed fishing tourism businesses have only reviewed destinations in limited locations. This is not surprising as almost half of users are “New Reviewers” as mentioned previously.
- **Explorer Badge:** This badge is assigned to users who are amongst the first to review a hotel, restaurant, or attraction in a given language. Our results indicate that 1 out of 3 users who have reviewed a fishing tourism business own this badge, meaning that they are trailblazers in the tourism domain, seeking out-of-the-beaten-path experiences.

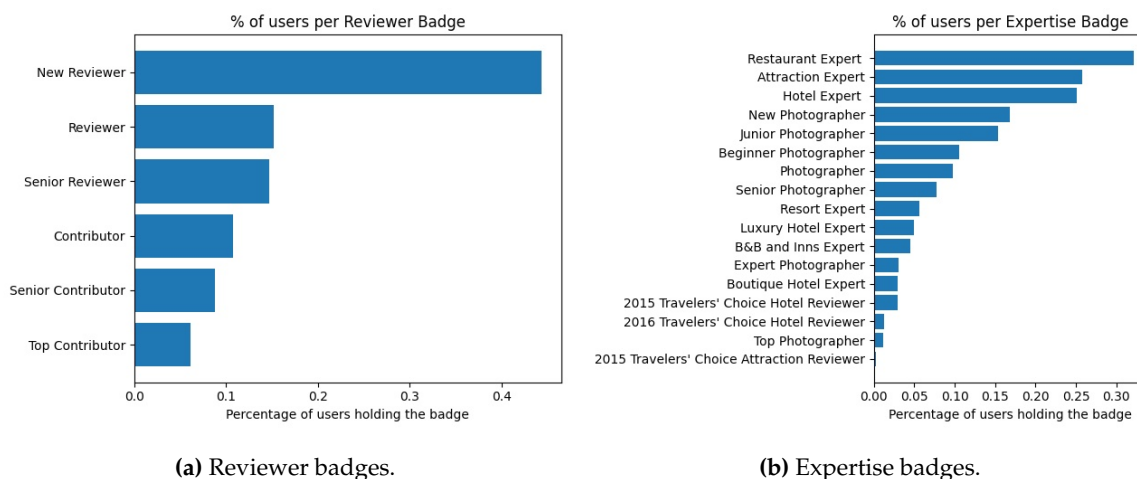


Figure 6. The distribution of different types of badges across all users who reviewed fishing tourism businesses (Source: Authors own creation).

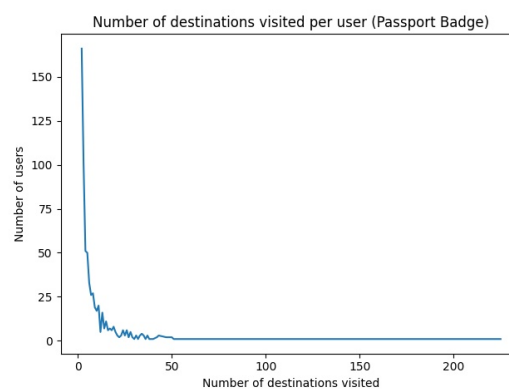


Figure 7. A plot of Passport badges indicating the number of locations visited across all users who reviewed fishing tourism businesses (Source: Authors own creation).

4.1.2. Sentiment Analysis and Emotion Extraction

Sentiment analysis refers to the opinion mining task that aims to discover the attitude of the author towards the discussed entity expressed in their texts. The sentiment is mainly classified as positive, neutral, and negative. Emotion extraction consists of a focused sentiment analysis task that

aims to extract specific emotions and not general attitudes. In our analysis, we employed the model of the six primary emotions defined by Ekman [27]: *joy, sadness, disgust, anger, fear, surprise*.

Sentiment analysis. To extract the total sentiment for each review, we employed a pre-trained model for sentiment analysis, the Textblob [28].

- **Insights for tourists.** The distribution of the overall sentiments found in user reviews in our corpus, reveals that there is a tendency to express neutral to positive comments about tourist venues and experiences. A surprising finding is that even for lower ratings, the detected sentiment polarity is rather positive than negative (see Figure 8a).
- **Insights for businesses.** The distribution of the sentiments found in reviews for fishing businesses, as shown in Figure 9, reflects the overall satisfaction of customers with the provided services. We expected a positive sentiment after the aforementioned linguistic insights and the high ratings fishing tourism businesses received.

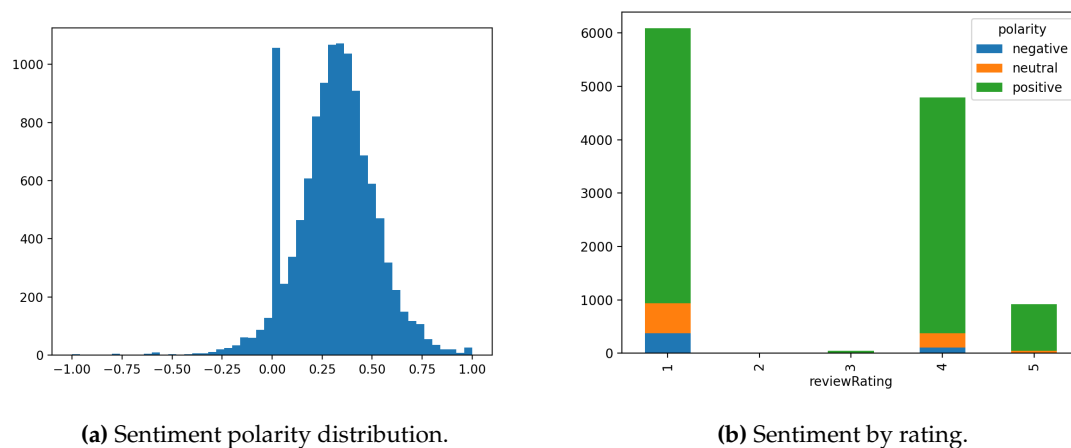


Figure 8. Overall users' sentiment polarity distribution and sentiment by rating for all reviews (Source: Authors own creation).

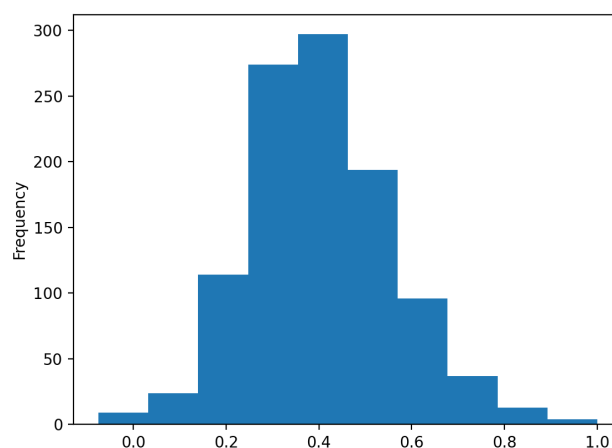


Figure 9. Fishing tourism businesses review sentiment polarity distribution (Source: Authors own creation).

Emotion extraction. To extract specific emotions expressed in reviews, we employed an unsupervised approach by employing an affective lexicon, including affective emojis. In particular, we

utilized “Wordnet Affect”, an extension of WordNet Domains, including a subset of synsets suitable to represent affective concepts correlated with affective words. We mapped words in reviews with the six primary emotions based on these affective concepts. If other specific emotions emerge, they are grouped into the general classes of positive and negative emotions. This approach works mainly for reviews written in English.

- **Insights for general tourist reviews.** We collected a set of 11,861 reviews from the profiles of users in our dataset. Given that a percentage of 90.8% of these reviews are written in English, we found 30,649 occurrences of affective concepts based on WordNet affect (see Table 2 for distributions). Emotion analysis shows that users are more likely to express positive emotions in their reviews, specifically surprise and joy, while negative emotions are expressed less frequently.

Table 2. Emotions found in users’ profiles (Source: Authors own creation).

Emotion	Percentage (counts)
positive emotion	56.5% (17,322)
negative emotion	4.8% (1,482)
other emotion	5.8% (1,798)
joy	7.3% (2,239)
surprise	20.4% (6,255)
anger	0.05% (176)
disgust	0.01% (32)
fear	1.1% (345)
sadness	3.2% (993)

- **Insights for fishing tourism businesses.** Concerning the affective concepts found in users’ reviews for fishing businesses, based on the linguistic and sentiment analysis preceded, we expect a high percentage of positive emotions in the total of 3,506 emotion terms found (see Table 3 for distributions). Indeed, positive emotions and surprise dominate, indicating customer satisfaction above expectation (prevalence of surprise emotion).

Table 3. Emotions found in user reviews for fishing businesses (Source: Authors own creation).

Emotion	Percentage (counts)
positive emotion	53.4% (1,875)
negative emotion	3.6% (128)
other emotion	5.0% (178)
joy	4.5% (161)
surprise	30.2% (1,061)
anger	0.1% (6)
disgust	0.2% (8)
fear	0.5% (21)
sadness	1.9% (68)

4.1.3. Topic Detection

Understanding a customer basis emerges from knowing customers’ needs and interests. This can be achieved by extracting topics and sub-topics of interests discussed or declared by users. For this task, we employed unsupervised learning techniques *k-means topic clustering* [29] and *Latent Dirichlet Allocation (LDA) topic modelling* [30]. Both algorithms define clusters of topics in a document collection given *k*, the predefined number of topics to be extracted. However, the main difference between the two methods is that *k-means* partitions the given documents into disjoint clusters (topics). At the same time, LDA assigns a document to a mixture of topics (one or more) with a representative percentage distribution [31]. We present the results at the collective fishing tourism business and user level to gain insights about the customer basis of the fishing tourism industry.

K-means topic clustering performed poorly in our data, as evaluated with elbow method and silhouette index. In particular, the elbow method reveals that the greater the number of k , the lower the sum of squared distances among clusters centroids, however, without converging to zero. Similar results are derived from silhouette index evaluation as the highest score achieved is lower than 0.5, indicating poor clustering, which was also evaluated manually. LDA typically requires a significant amount of text to identify well-defined topics. Despite the relatively small volume of reviews in our dataset, LDA-emerged topics seem more reasonable than the k-means approach. In our case, after manual experimentation and human judgment, we identified four distinct super topics, as shown in Figure 10: hotel reviews and accommodation, food experience, general service quality, and fishing experiences. This reflects that fishing is a primary and general interest of users who visit fishing boat businesses. Another interesting insight regarding the third topic (fishing tourism) is that the included reviews are more personal than average since the most relevant term for this topic is actually the boat's owner's name. Regarding the remaining topics, it is worth mentioning that the personnel of a touristic entity is amongst the most relevant terms in both the first and second topics. This signifies the high importance of proper staffing, especially in the hotels and accommodation industry, where the term "staff" appears even higher than the hotel's amenities (rooms, pool, bar, restaurant, etc.).



Figure 10. Intertopic distance map (via multidimensional scaling) and top-30 most relevant terms for each topic (Source: Authors own creation).

4.2. User Profiling Aspects: Gender, Age and Marital Status

Demographic information of customers is a crucial aspect for any business to identify its audience and their specific needs and customize its products or/and services to maximize consumer satisfaction. Based on the information provided on users' profiles, we were able to extract the explicitly declared information about their gender, age, and marital status. With these ground truth data for a subset of our dataset, we train classifiers to infer this demographic information for the rest of the users.

4.2.1. Gender Classification

Following the generic methodology presented in Figure 11, we describe the steps and the obtained results for the gender prediction task.

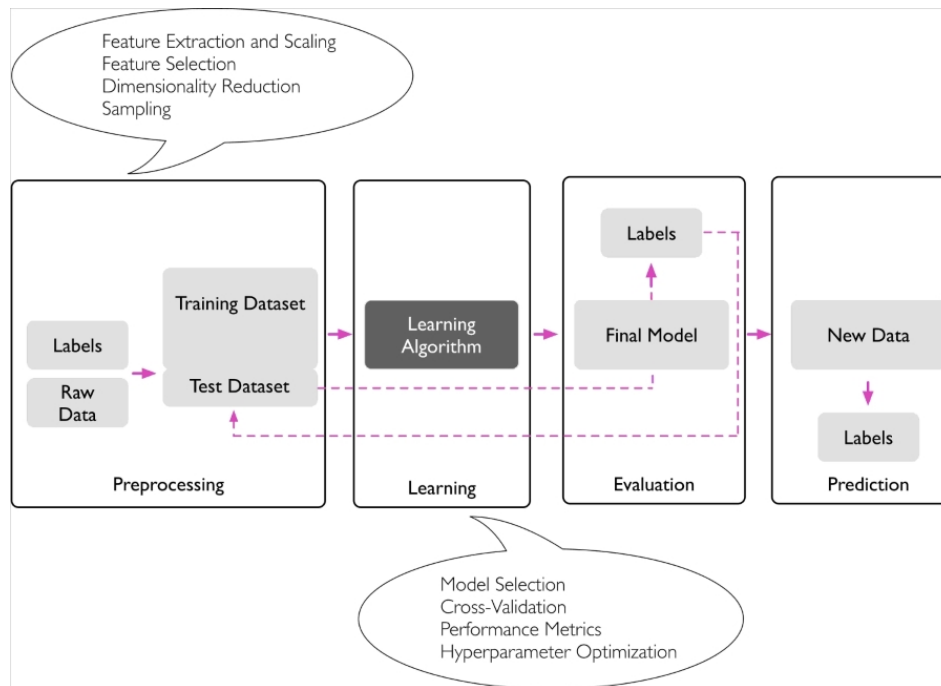


Figure 11. Typical machine learning process for predictive modeling (Figure courtesy of Raschka and Mirjalili [32]).

Ground truth. Individuals' gender in our dataset could take one out of two labels: woman or man. Thus, our challenge is to build a binary classification model. As seen in Figure 12, in our dataset, 28.2% of our users are women, 18.4% are men, and 53.3% have no label available (nan). For this subset, we build the gender classifier to infer their gender.

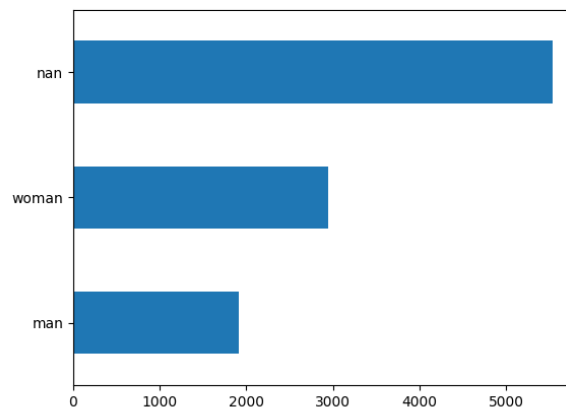


Figure 12. Gender labels in our dataset (Source: Authors own creation).

Features extraction. In order to train our gender classifier, we need to extract the most appropriate features arising from related studies [33,34]. These can be summarized as follows:

- **Gender estimation based on name:** We train a naive Bayes model in order to get an estimation of each user's gender based on their username according to names lexicons with female and male names. This is assumed a powerful feature for gender detection.
- **Sentiment score:** The aggregated score of reviews sentiment.
- **Syntax features:** Number of part of speech tags (Adjectives, Nouns, Verbs, and Adverbs).

- **Language vectors:** The frequency vectors of words used by each user. We constructed TF-IDF and n-grams vectors reflecting the importance of words and phrases in a collection of documents based on preprocessed textual data.

Model training. We experimented with different types of classifiers, as mentioned above, such as logistic regression (lr), random forest (rf), and stochastic gradient descent (sdg). We split the ground truth to train and test sets while keeping 70% for training and 30% for testing each model. We followed cross-validation with 10 iterations in order to check model accuracy. We evaluate the performance of our models with accuracy and f1 score. The results are presented in Table 4.

Table 4. Gender classifier training (Source: Authors own creation).

Classifier		Training error	Test error
lr	acc.	0.84	0.73
	f1	0.83	0.72
rf	acc.	1.0	0.71
	f1	1.0	0.68
sgd	acc.	0.40	0.42
	f1	0.23	0.25

The best results are achieved with the logistic regression classifier achieving an accuracy score of 73%. In specific, the model is confused and is most likely to predict the man label falsely. After training, we applied the model to our dataset to annotate it and obtain labels for the nan cases. The gender-labeled data are used to extract insights about the 4Es.

4.2.2. Age Classifier

Following the same methodology, we describe the steps and the obtained results for the age prediction task.

Ground truth. Individuals' age exists in a total of 32% in our dataset and falls in one of the classes 18-24 (0.25%), 25-34 (4.6%), 35-49 (14.9%), 50-64 (12.3%). In Figure 13, we noticed that there is a high imbalance in class frequency, and as a result, to build a valid classifier, we need to apply class imbalance machine learning techniques.

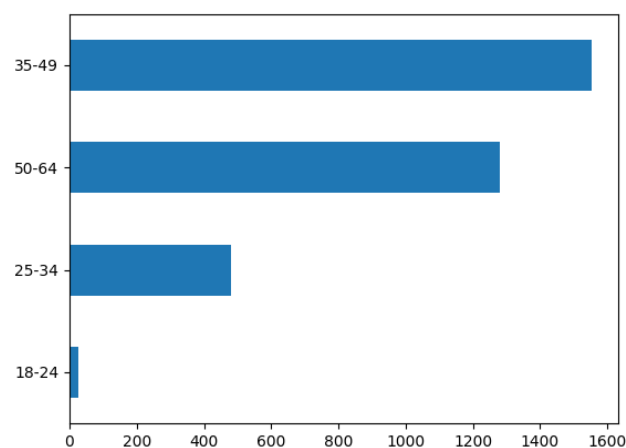


Figure 13. Age labels in our dataset (Source: Authors own creation).

Features extraction. Similarly, with the case of the gender classifier, we need to extract the most appropriate features arising from related studies [34] to build the age classifier. These can be summarized as follows:

- **Structure features:** Refer to the structural use of language. These features include the number of words in each review, the number of characters, the number of words in a sentence, and the number of exclamation marks.
- **Syntax features:** Refer to the number of parts of speech tags (Adjectives, Nouns, Verbs, and Adverbs).
- **Sentiment score:** The aggregated score of reviews sentiment.
- **Readability features:** Refer to the level of the text complexity. We included: a) Flesch reading ease, indicating how easy is a text to read, b) Smog index, estimating the years of education needed to understand a piece of writing, c) Flesch–Kincaid grade, indicating the average student in that grade level can read the text, d) Coleman Liau index, gauging the understandability of a text, e) automated readability index, assessing the understandability of a text, f) Dale Chall readability score, providing a numeric gauge of the comprehension difficulty that readers come upon when reading a text, g) difficult words, indicating how many difficult words used in a text, h) gunning fog, estimating the years of formal education a person needs to understand the text on the first reading.
- **Language vectors:** The frequency vectors of words used by each user. We constructed TF-IDF and n-grams vectors reflecting the importance of words and phrases in a collection of documents based on preprocessed textual data.

Imbalance learning. As already mentioned, our ground-truth dataset is imbalanced. We experiment with different techniques to handle this challenge as oversampling and undersampling, use of synthetic examples, etc, and we conclude that the Synthetic Minority Oversampling Technique (SMOTE), which synthesizes new examples for the minority class, gives the best results. Specifically, SMOTE works by selecting examples that are close to the feature space, drawing a line between the examples in the feature space, and drawing a new sample at a point along that line. A random example from the minority class is first chosen. Then k of the nearest neighbors for that example is found (typically $k=5$). A randomly selected neighbor is chosen, and a synthetic example is created at a randomly selected point between the two examples in feature space [35]. We present our results with and without the use of SMOTE technique. Our sampling strategy follows the rule of oversampling the minority class up to 10% of the majority class.

Model training. We experimented with different types of classifiers as mentioned above, such as logistic regression (lr), random forest (rf), and stochastic gradient descent (sgd). We split the ground truth to train and test sets while keeping 70% for training and 30% for testing each model. We followed cross-validation with 10 iterations in order to check the model's accuracy. We evaluate the performance of our models with accuracy and f1 score. The results are present in Table 5.

Table 5. Age classifier training (Source: Authors own creation).

Classifier		Training error	Test error
lr	acc.	0.47	0.46
	f1	0.42	0.42
rf	acc.	1.0	0.62
	f1	1.0	0.59
sgd	acc.	0.44	0.46
	f1	0.32	0.34

Best results are achieved with random forest classifier achieving an accuracy score of 62%. In specific, the model is confused and is most likely to falsely predict the age categories 35-49 and 50-64. After training, we applied the model to our dataset to annotate it and obtain labels for the unlabelled cases. The age-labeled data are used to extract insights about the 4Es.

4.2.3. Marital Status Detection

Extracting the marital status of each reviewer is a detection task since we do not have a ground truth to train a machine learning model and infer the status of the travelers in our dataset. As a result, based on a lexicon approach, we define vocabulary sets to detect if the reviewer uses words or phrases that indicate if they travel with their *family* or their *partner*. In the cases where there is not any relevant info available, we assign the class *unknown*.

We define the marital classes as follows: a) family travelers (those who travel with kids): use of words like children, child, kid(s), son, daughter, dad, father, mom, mum, and mother. b) couple travelers (those who travel with their partner, no kids): use of words like husband, wife, spouse, girlfriend, boyfriend, partner. c) unknown, for the rest of the cases, they can be solo travelers, a group of people traveling together, or any other case.

As seen in Figure 14, we detect family travelers at 22.9%, couple travelers at 7.3%, and the unknown class at 70%. We present the tourism preferences of different marital statuses as well as the relationship with demographics and 4E classes.

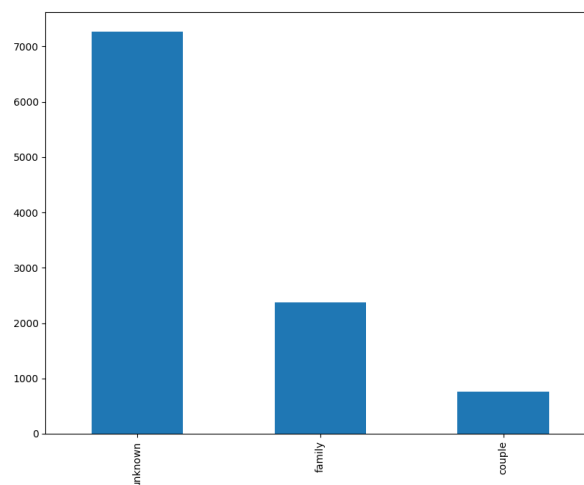


Figure 14. Marital status distribution (Source: Authors own creation).

4.3. The 4Es of Experience Economy

In this section, we dive into the different experience realms of the fishing tourism industry as seen through the lens of TripAdvisor tourist reviews. To acquire information regarding the experience dimensions, i.e., “Educational”, “Entertainment”, “Aesthetic”, “Escapist”, expressed in each review, we manually annotated 240 user reviews. In other words, human annotators read and annotated a number of reviews each, according to the four dimensions of the Experience Economy.

The heat map in Figure 15 visualized the results of this annotation process in the form of a frequency matrix for the four dimensions. We notice that “Entertainment” is by far the prevalent dimension in the tourists’ reviews (88% of reviews), followed by “Aesthetic” (34% of reviews), “Educational” (32% of reviews), and “Escapist” (19% of reviews). Not surprisingly, when it comes down to the most prevalent pairs of dimensions, the pairs of “Entertainment”-“Aesthetic” and “Educational”-“Entertainment” are the most frequently, co-existing in 30% and 28% of the reviews, respectively. However, to get a better idea of the true co-existence of dimensions that is not biased by the frequency of appearance of the individual experience realms, we visualize the Jaccard Similarity index for each dimension pair in Figure 16. The Jaccard similarity index “compares members for two sets to see which members are shared and which are distinct”. It’s “a measure of similarity for the two sets of

data, with a range from 0 to 1⁴. The higher the percentage, the more similar the two populations are. In the case of tourist reviews, the higher the Jaccard index, the more reviews are shared among the two dimensions. Now, apart from the pairs, we discussed previously, another common pair emerges, specifically the “Aesthetic”-“Escapist” combination (Jaccard index of 0.26), denoting the co-existence of the “Aesthetic” and “Escapist” experiences in the fishing tourism reviews.

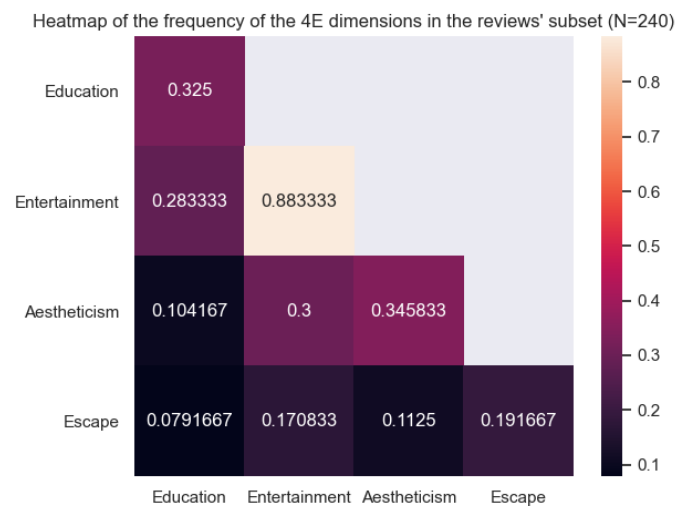


Figure 15. Frequency matrix for the four dimensions of the Experience Economy (Source: Authors own creation).

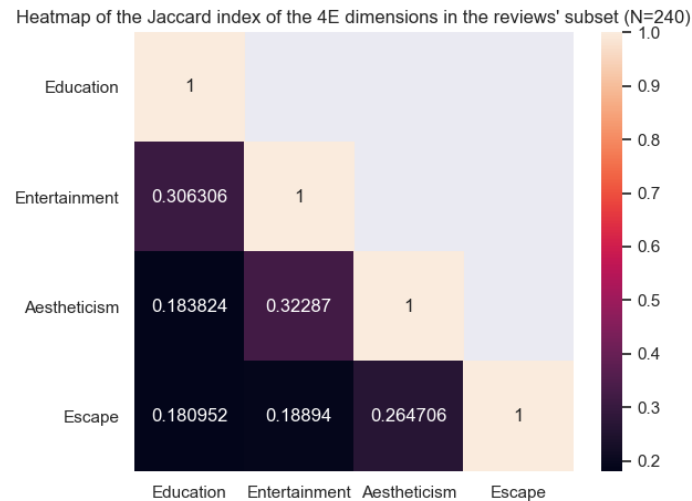


Figure 16. Jaccard Similarity matrix for the four dimensions of the Experience Economy (Source: Authors own creation).

With regards to the demographics of the users within each experience realm, we visualize the age, gender, and marital status of the reviewers in Figures 17–19, respectively. Regarding age, the distribution of age groups appearing in the different experience realms does not vary significantly. There is only a slight difference in the Entertainment realm, where the oldest age group (50-64) has a larger percentage compared to the other realms, potentially declaring that fishing tourism is

⁴ <https://www.statisticshowto.com/jaccard-index/>

particularly enjoyable for older adults. Similarly, the age distribution does not vary significantly between the different dimensions of the Experience Economy, with most reviews coming from female users, as seen in Figure 18. Lastly, concerning marital status, the percentage of reviews coming from couples is almost steady across all dimensions. However, when it comes to family users, the “Aesthetic” and “Escapist” dimensions show higher percentages, meaning that these experiences are possibly felt more intensely by this user group.

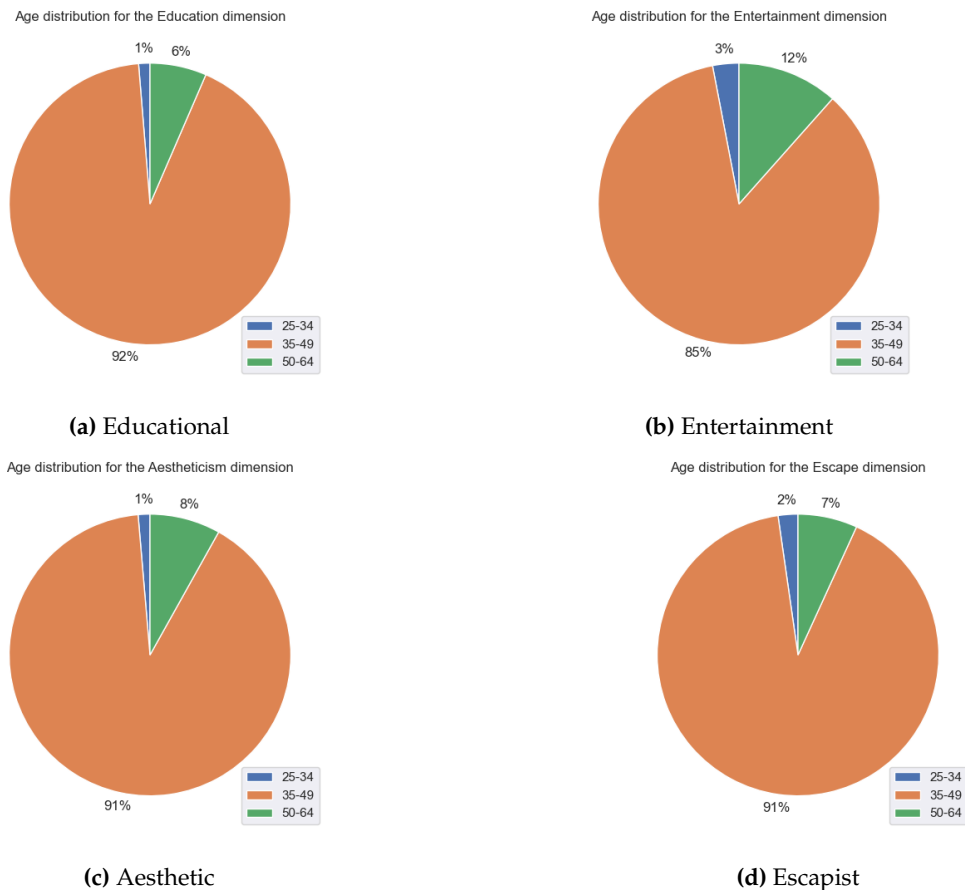


Figure 17. The age distribution of reviewers per experience realm (Source: Authors own creation).

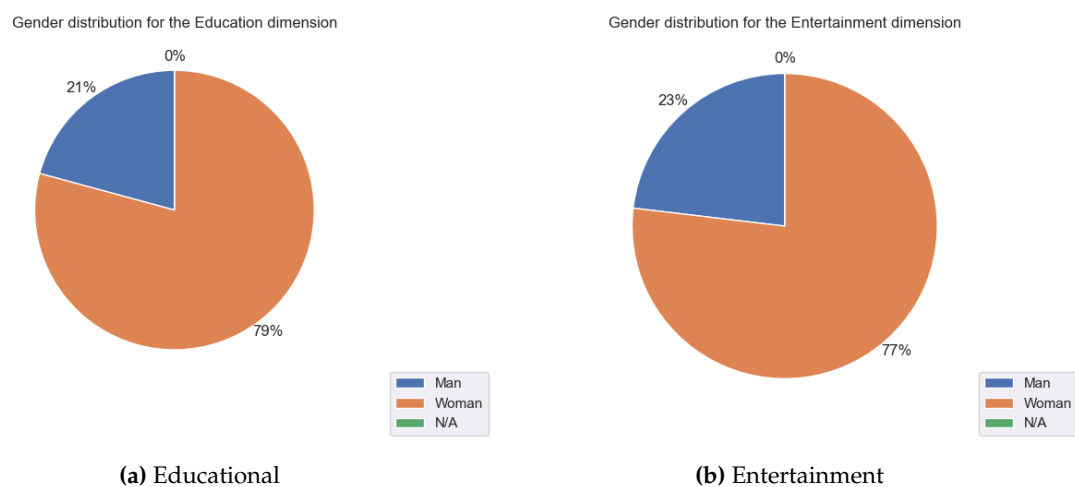


Figure 18. Cont.

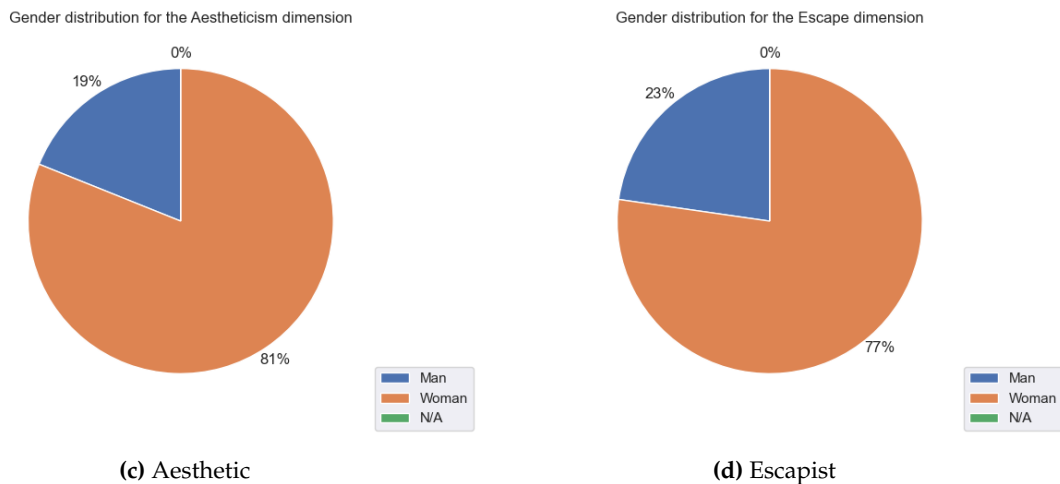


Figure 18. The gender distribution of reviewers per experience realm (Source: Authors own creation).

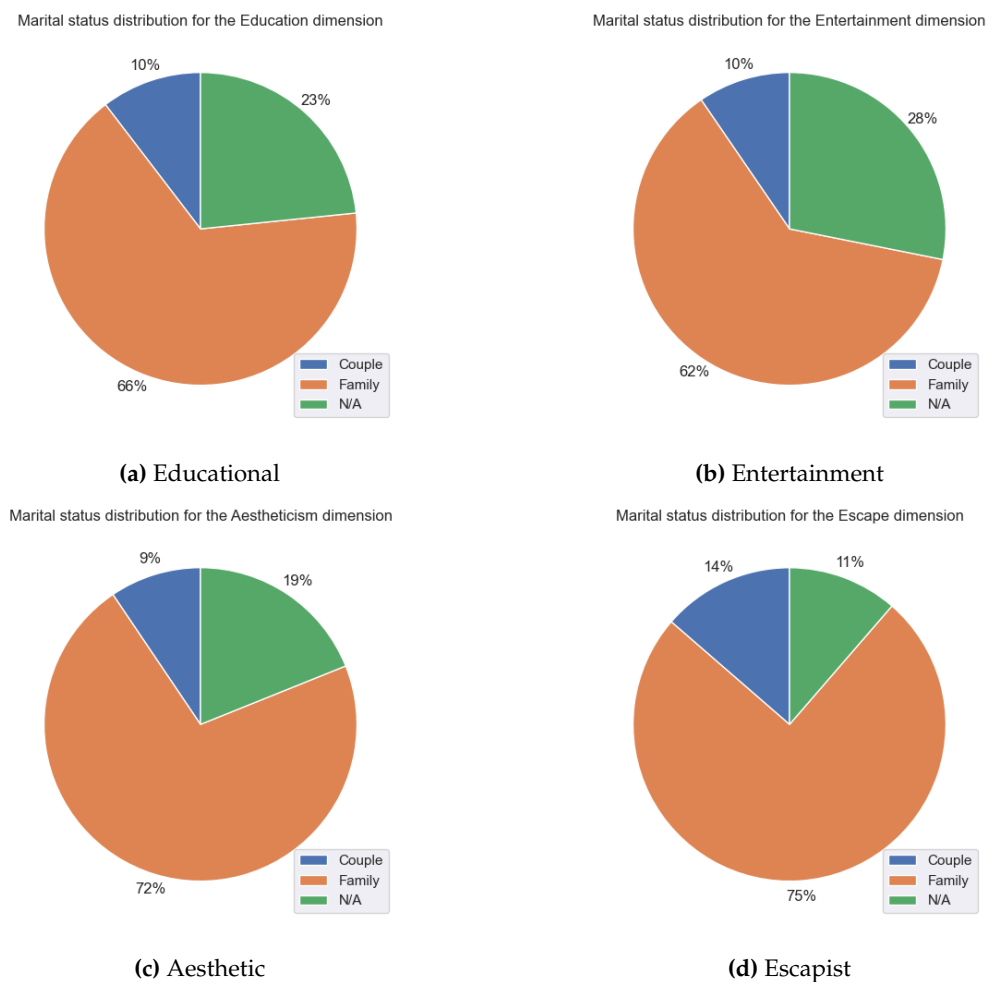


Figure 19. The marital distribution of reviewers per experience realm (Source: Authors own creation).

Finally, the linguistic exploration of reviews belonging to different realms did not produce any significant results.

5. Discussion & Conclusions

Tourists went on boat trips mostly for fishing and actively participated in the process (fishing experience, caught fish, etc). The overall experience of fishing boat trips is highly recommended as

tourists mention that they had a “great time” and “fantastic day”. Tourists appreciate the beauty of the natural environment by leaving positive comments about the “crystal waters”, the sea, the fresh fish, etc. They highlight the hospitality and the skills of the crew and business owners. These reviews reflect the positive tourist experience translated into star ratings from the tourists. It’s evident that the vast majority of tourists left 5-star rated reviews. Tourists overall put emphasis on commenting on the offered services. Emotion analysis shows that users are more likely to express positive emotions in their reviews, specifically surprise and joy, while negative emotions are expressed less frequently. Indeed, positive emotions and surprise dominate, indicating customer satisfaction above expectation.

Interestingly, approximately half the users belong to the “New Reviewer” category, meaning that they only joined the TripAdvisor platform to positively review the respective fishing tourism business. The users who review fishing tourism businesses also tend to review hotels and attractions, as well as luxury and boutique hotels, and B&B and Inns at a smaller scale. Fishing is a primary and general interest of users who visit fishing boat businesses. Another interesting insight regarding fishing tourism is that the included reviews are more personal than average since the most relevant term for this topic is actually the boat’s owner’s name.

“Entertainment” is by far the prevalent dimension in the tourists’ reviews, followed by “Aesthetic”, “Educational” and “Escapist”. Not surprisingly, when it comes down to the most prevalent pairs of dimensions, the pairs of “Entertainment”-“Aesthetic” and “Educational”-“Entertainment” are the most frequent. Another common pair emerges, specifically the “Aesthetic”-“Escapist” combination, denoting the co-existence of the “Aesthetic” and “Escapist” experiences in the fishing tourism reviews. To truly transfer Value as a business or organization, we need to understand and act on the perceptions of customers about Quality and Value, the process of creating that Value, and efficient and effective management of the same time resources in order to create this Value (Grönroos, 2000).

Nowadays, fishing trip organizers list on specialized websites the fishing package or packages they intend to offer, i.e., detailing what they will provide to the client (tourist), in which areas, and at what cost. The client can make an on-the-spot online booking and payment, and the fisherman will be informed immediately about the booking by email or SMS. Many fishing trip organizers have already signed contracts, and several are in the process of signing. These websites also cover tourists who are interested not only in fishing but also in being shown a way of fishing or even being trained in these types of fishing. In particular, in Greece, all the forms of fishing that are possible have been included: a) fishing from boats, b) kayak fishing, c) fishing from the shore of all types (Casting, English, Spinning, etc.), d) fishing in lakes and rivers and e) spearfishing. Each new tourist activity goes through a period of evolution and adaptation within the community, earning the participants’ loyalty and reinforcing repeated tourism, which contributes to the development and improvement of the local tourism market. Fishing tourism as a new tourist activity provides potential benefits from fishing tourism to local communities. Therefore, the necessary attention should be given to developing a strategy in this tourism sector. This development will bring new opportunities and challenges, it can offer professional opportunities to fishermen and will increase the rural economy if governments provide adequate infrastructure, leadership, legislative and financial support that will set the foundation for sustainable development in long terms [36]. The combination of fishing tourism and marine reserves emerges as the optimal strategy, and that the presence of visitors in these areas generates larger profits than if only fishing was considered [37].

Theoretical implications on fishing tourism can encompass a wide range of considerations that relate to the interactions between tourism and fishing activities. These implications may arise from various academic disciplines, including economics, ecology, sociology, and environmental science. Here are some theoretical implications to consider: the combination of the dimensions and the opportunity of developing value on recreational fishing tourism. More over there are many impacts which are now open for research in various areas. Economic impact: Fishing tourism can have multiplier effects on local economies. Tourists’ expenditures on accommodations, equipment rental, guides, and other services can generate indirect and induced economic impacts in the host community.

Ecological Impact: Fishing tourism can lead to overfishing if not managed sustainably. The theoretical implications involve the need for effective regulations and strategies to prevent overexploitation of fish populations and protect marine ecosystems. **Ecotourism:** Sustainable fishing tourism models can contribute to the conservation of natural resources by fostering awareness and appreciation for aquatic ecosystems. **Sociocultural Impact:**

- **Cultural Exchange:** Fishing tourism can facilitate cultural exchange between tourists and local fishing communities, leading to mutual understanding and preservation of traditional practices. This implies the importance of promoting respectful interactions and cultural sensitivity.
- **Community Livelihoods:** The theoretical implications involve examining how fishing tourism affects the livelihoods and well-being of local communities. Sustainable fishing tourism can enhance income diversification and improve quality of life.
- **Tourism Management:** Theoretical discussions revolve around determining the carrying capacity of fishing tourism destinations to ensure that environmental and social impacts are kept within sustainable limits.
- **Stakeholder Engagement:** Effective stakeholder engagement is crucial for managing fishing tourism. The implications include the need for collaboration among governments, local communities, tour operators, and conservation organizations.
- **Environmental Ethics:** Theoretical implications extend to ethical discussions about catch-and-release practices, the welfare of targeted fish species, and the broader ecological consequences of fishing tourism.
- **Conservation and Research:** Fishing tourism can provide opportunities for scientific research, such as studying fish populations, migration patterns, and ecosystem dynamics. Theoretical implications emphasize the role of fishing tourism in advancing marine conservation efforts.
- **Education and Outreach:** Fishing tourism can serve as a platform for educating tourists and the public about the importance of marine conservation, fostering a sense of responsibility and support for preserving aquatic ecosystems.
- **Climate Change Adaptation:** Theoretical implications may explore how fishing tourism destinations need to adapt to changing climate conditions, such as shifts in fish distribution and abundance, and how these changes could affect tourism experiences and local economies.

It is important to note that the actual implications of fishing tourism will depend on factors such as destination characteristics, management strategies, regulatory frameworks, and the behaviors of tourists and local communities. Researchers and practitioners in various fields continue to explore these implications to ensure that fishing tourism contributes positively to both the environment and the well-being of host communities.

Author Contributions: Conceptualization, Y.G. and Y.S.; methodology, Y.S., Y.G., and K.D.; software, Y.S. and K.D.; validation, Y.G., Y.S., M.M. and L.A.; formal analysis, Y.G., Y.S., M.M. and L.A.; investigation, Y.G., Y.S., M.M., V.V. and L.A.; resources, Y.G. and B.P.; data curation, Y.S. and K.D.; writing—original draft preparation, Y.G., Y.S. and K.D.; writing—review and editing, Y.G., B.P. and P.A.; visualization, Y.S. and K.D.; supervision, Y.G. and Y.S.; project administration, Y.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable, as the study did not involve any data acquired through intervention or interaction with individual human subjects, or any identifiable private information—all data collected were anonymous.

Data Availability Statement: Data are unavailable due to privacy restrictions.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

API	Application Programming Interface
DUTH	Democritus University of Thrace
HCMR	Hellenic Centre for Marine Research
HTML	HyperText Markup Language
LDA	Latent Dirichlet Allocation
NLP	Natural Language Processing
POS	Part Of Speech
SMOTE	Synthetic Minority Oversampling Technique
TF-IDF	Term Frequency - Inverse Document Frequency
XML	Extensible Markup Language

References

1. Jafari, J. *Encyclopedia of tourism*; Routledge, 2002.
2. Centre for the Promotion of Imports from developing countries. The European market potential for sports tourism. *Ministry of Foreign Affairs of the Netherlands* 2021.
3. Pine, B.J.; Pine, J.; Gilmore, J.H. *The experience economy: work is theatre & every business a stage*; Harvard Business Press, 1999.
4. Ali-Knight, J. The role of niche tourism products in destination development. PhD thesis, Edinburgh Napier University, 2010.
5. Holbrook, M.B.; Hirschman, E.C. The experiential aspects of consumption: Consumer fantasies, feelings, and fun. *Journal of consumer research* 1982, 9, 132–140.
6. Pine, B.J.; Gilmore, J.H.; et al. *Welcome to the experience economy*; Vol. 76, Harvard Business Review Press Cambridge, MA, USA, 1998.
7. Pikkemaat, B.; Weiermair, K. The aesthetic (design) orientated customer in tourism-implications for product development. In Proceedings of the EIASM 10th international product development management conference. EIASM Brussels, 2003, Vol. 825, p. 839.
8. Menner, T.; Höpken, W.; Fuchs, M.; Lexhagen, M. Topic detection: identifying relevant topics in tourism reviews. In *Information and Communication Technologies in Tourism 2016*; Springer, 2016; pp. 411–423.
9. Rossetti, M.; Stella, F.; Zanker, M. Analyzing user reviews in tourism with topic models. *Information Technology & Tourism* 2016, 16, 5–21.
10. Afzaal, M.; Usman, M.; Fong, A.C.; Fong, S. Multiaspect-based opinion classification model for tourist reviews. *Expert Systems* 2019, 36, e12371.
11. Afzaal, M.; Usman, M.; Fong, A. Predictive aspect-based sentiment classification of online tourist reviews. *Journal of Information Science* 2019, 45, 341–363.
12. Yu, C.; Zhu, X.; Feng, B.; Cai, L.; An, L. Sentiment Analysis of Japanese Tourism Online Reviews. *Journal of Data and Information Science* 2019, 4, 89–113.
13. Marrese-Taylor, E.; Velásquez, J.D.; Bravo-Marquez, F. A novel deterministic approach for aspect-based opinion mining in tourism products reviews. *Expert Systems with Applications* 2014, 41, 7764–7775.
14. Kirilenko, A.P.; Stepchenkova, S.O.; Kim, H.; Li, X. Automated sentiment analysis in tourism: Comparison of approaches. *Journal of Travel Research* 2018, 57, 1012–1025.
15. Fang, B.; Ye, Q.; Kucukusta, D.; Law, R. Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. *Tourism Management* 2016, 52, 498–506.
16. Shin, S.; Du, Q.; Xiang, Z. What's vs. how's in online hotel reviews: Comparing information value of content and writing style with machine learning. In *Information and Communication Technologies in Tourism 2019*; Springer, 2019; pp. 321–332.
17. Leal, F.; González-Vélez, H.; Malheiro, B.; Burguillo, J.C. Semantic profiling and destination recommendation based on crowd-sourced tourist reviews. In Proceedings of the International Symposium on Distributed Computing and Artificial Intelligence. Springer, 2017, pp. 140–147.
18. Guerreiro, J.; Rita, P. How to predict explicit recommendations in online reviews using text mining and sentiment analysis. *Journal of Hospitality and Tourism Management* 2020, 43, 269–272.
19. Dong, X.; Li, T.; Song, R.; Ding, Z. Profiling users via their reviews: an extended systematic mapping study. *Software and Systems Modeling* 2020, pp. 1–21.

20. Madke, N.B.; Kulkarni, M.Y.; Mule, M.R.; Lakade, M.A.; Kulkarni, M.S. User Profile Based Behavior Identificaton Using Data Mining Technique **2018**.
21. Farnadi, G.; Tang, J.; De Cock, M.; Moens, M.F. User profiling through deep multimodal fusion. In Proceedings of the Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, 2018, pp. 171–179.
22. Alkan, O.; Daly, E. User Profiling from Reviews for Accurate Time-Based Recommendations. *arXiv preprint arXiv:2006.08805* **2020**.
23. Kavitha, S.; Jobi, V.; Rajeswari, S. Tourism recommendation using social media profiles. In *Artificial Intelligence and Evolutionary Computations in Engineering Systems*; Springer, 2017; pp. 243–253.
24. Missaoui, S.; Viviani, M.; Faiz, R.; Pasi, G. A language modeling approach for the recommendation of tourism-related services. In Proceedings of the Proceedings of the Symposium on Applied Computing, 2017, pp. 1697–1700.
25. Fazzolari, M.; Petrocchi, M. A study on online travel reviews through intelligent data analysis. *Information Technology & Tourism* **2018**, *20*, 37–58.
26. Jivani, A.G.; et al. A comparative study of stemming algorithms. *Int. J. Comp. Tech. Appl* **2011**, *2*, 1930–1938.
27. Ekman, P.; Sorenson, E.R.; Friesen, W.V. Pan-cultural elements in facial displays of emotion. *Science* **1969**, *164*, 86–88.
28. Loria, S. textblob Documentation. *Release 0.15* **2018**, *2*.
29. Likas, A.; Vlassis, N.; Verbeek, J.J. The global k-means clustering algorithm. *Pattern recognition* **2003**, *36*, 451–461.
30. Ramage, D.; Hall, D.; Nallapati, R.; Manning, C.D. Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. In Proceedings of the Proceedings of the 2009 conference on empirical methods in natural language processing, 2009, pp. 248–256.
31. Kelaiaia, A.; Merouani, H.F. Clustering with probabilistic topic models on arabic texts: a comparative study of LDA and K-means. *Int. Arab J. Inf. Technol.* **2016**, *13*, 332–338.
32. Raschka, S.; Mirjalili, V. Python machine learning second edition, 2017.
33. Marquardt, J.; Farnadi, G.; Vasudevan, G.; Moens, M.F.; Davalos, S.; Teredesai, A.; De Cock, M. Age and gender identification in social media. *Proceedings of CLEF 2014 Evaluation Labs* **2014**, *1180*, 1129–1136.
34. Weren, E.R.; Kauer, A.U.; Mizusaki, L.; Moreira, V.P.; de Oliveira, J.P.M.; Wives, L.K. Examining multiple features for author profiling. *Journal of information and data management* **2014**, *5*, 266–266.
35. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research* **2002**, *16*, 321–357.
36. Yfantidou, G.; Matarazzo, M. The future of sustainable tourism in developing countries. *Sustainable development* **2017**, *25*, 459–466.
37. Falcó, C.; Moeller, H.V. Optimal spatial management in a multiuse marine habitat: Balancing fisheries and tourism. *Natural Resource Modeling* **2022**, *35*, e12309.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.