Article

CSR and Corporate Sustainability: Theoretical and Empirical Approaches Based on Data Science in Spanish Tourism Companies

Maria Fernanda Bernal Salazar *[ID], **Elisa Baraibar-Diez** [ID] **and Jesús Collado-Agudo** [ID]

Department of Business Administration, Faculty of Economics and Business Administration, University of Cantabria, 39005 Santander, Spain; elisa.baraibar@unican.es (E.B.-D.); jesus.collado@unican.es (J.C.-A.)
* Correspondence: maria-fernanda.bernal@alumnos.unican.es

Abstract: This study combines a theoretical and empirical approach to analyze the transition from corporate social responsibility to corporate sustainability in Spanish tourism companies, with an emphasis on the integration of ESG (environmental, social, and governance) criteria. In the theoretical domain, a computational literature review is conducted by applying topic modeling to 1505 scientific documents published between 2004 and 2023, identifying key trends and evaluating the evolution from CSR to CS. In the empirical domain, 364 corporate reports published between 2010 and 2021 are analyzed, using text mining techniques to examine changes in the relative frequency of terms associated with CSR and CS, and the BERTopic model to detect key management areas. Additionally, the FinBERT model classifies the content of the reports into nine ESG categories, quantifying their relevance across different tourism subsectors. The results confirm a progressive transition towards CS, evidenced by shifts in thematic priorities reflected in the literature and a significant increase in the use of terms associated with CS in corporate reports. The research provides valuable insights for managers, regulators, and local communities, enabling the design of strategies better aligned with ESG standards, optimizing business management, and strengthening sustainability in the Spanish tourism sector.

Keywords: corporate sustainability; corporate social responsibility; text mining; BERTopic; FinBERT; ESG; corporate reporting; tourism

1. Introduction

In the tourism sector, corporate social responsibility (CSR) and corporate sustainability (CS) are fundamental for companies to achieve their economic objectives while effectively responding to social and environmental demands, contributing to sustainable development [1–3]. Although both concepts aim to balance economic, social, and environmental dimensions, CS has emerged as a more holistic and integrated approach oriented towards long-term sustainability [4,5]. The transition from CSR to CS has been driven by international frameworks such as ESG (environmental, social, and governance) criteria, which promote the integration of sustainability and ethical practices across all aspects of business management, enabling their quantitative and strategic evaluation [6]. According to Back [7], the ESG approach focuses on managing and developing sustainability, a process accelerated by global initiatives such as the 2030 Agenda and the United Nations Sustainable Development Goals (SDGs) in 2015, as well as the transformative impact of the COVID-19 pandemic, which highlighted the need for resilient and sustainable business strategies.

The tourism sector's reliance on natural and human resources makes ESG practices essential for optimizing waste management, fostering community engagement, and promoting equitable working conditions [7]. Consequently, ESG has become one of the most common ways to implement and materialize CS. However, this evolution faces significant challenges, both conceptual and practical, that still limit its full adoption [8]. Moreover, stakeholders demand that companies not only operate responsibly but also clearly communicate their economic, social, and environmental outcomes [9,10]. Although ESG reports have evolved into integrated formats combining financial and non-financial information [11], gaps persist in areas such as biodiversity conservation and renewable energy integration [12].

In Spain, sustainability is a strategic pillar of the tourism sector. The country is recognized as one of the most attractive tourist destinations for travelers committed to sustainability. A total of 38% of sustainable tourists view Spain favorably, and 97% of visitors report high levels of satisfaction with their experience [13]. However, the projected growth of tourism underscores the need to balance economic benefits with potential negative social and environmental impacts [14]. This context highlights the importance of analyzing how tourism companies are adopting and communicating CSR and CS practices in their corporate reports.

This research addresses these issues by combining theoretical and empirical approaches to analyze the perception and evolution of CSR and CS in tourism. In the theoretical domain, a computational literature review is conducted between 2004 and 2023, employing topic modeling techniques to identify major research trends and statistically validate the conceptual evolution of CSR and CS. In the empirical domain, 364 corporate reports published by Spanish tourism companies between 2010 and 2021 are analyzed. This analysis includes applying the BERTopic model to identify key management areas in CSR and CS, as well as text mining techniques to evaluate the frequency and relevance of terms related to these themes over time. Additionally, using the natural language processing (NLP) model FinBERT, the textual content of corporate reports is examined, quantifying the importance that different subsectors (transport, hotels and accommodations, and restaurants) attribute to specific ESG aspects, such as climate change, natural capital, pollution and waste, human capital, product responsibility, community relations, corporate governance, ethics, and corporate values.

Specifically, this research makes the following contributions:

- It provides an updated and detailed view of the theoretical fields of CSR and CS in tourism, evaluating the shift in research focus from CSR to CS. In the context of sustainable tourism, this is essential because a solid theoretical understanding can promote coherent changes that reduce ambiguity in academia and the implementation of business practices [15].
- It introduces an innovative methodology that combines data science techniques for analyzing sustainability in tourism which can be replicated in future studies. This enhances the focus of tourism research, as studies applying data science and big data analysis are limited [16–18].
- It increases scientific knowledge about how Spanish tourism companies approach the theory of CSR and CS. Recent studies, such as those by Koh [19] and Back [7], highlight the need for more longitudinal and sectoral studies. These are useful for evaluating evolution over time and identifying the importance different subsectors attribute to sustainability.

2. Literature Review

2.1. Evolution of CSR and CS in Tourism

CSR has historically represented a business strategy for managing social, economic, and environmental aspects [4,20]. In the tourism sector, academic interest has grown significantly, especially since 2018, experiencing an annual growth rate of 13.9%, highlighting its importance in the industry [3,21,22]. Studies have explored a wide range of topics, including environmental management, financial performance, marketing perspectives, tourist behavior, customer loyalty, and corporate reputation [2,23–27]. According to Yoopetch et al. [28], research on CSR in tourism can be grouped into four main areas: CSR performance or outcomes, CSR policy, business ethics, and sustainable tourism and the environment. Meanwhile, Sánchez-Camacho et al. [24] identify eight key areas, i.e., business activity and financial performance; attitude and behavior; marketing perspective; sustainable tourism; sustainability; government action; environmental awareness and motivation; and the impact of major events.

CSR has provided companies with competitive advantages, positively contributing to the environment, employee well-being, customer satisfaction, and community development [29]. However, in recent years, a paradigm shift toward CS has emerged. This approach seeks to create long-term value for stakeholders by comprehensively managing opportunities and risks across economic, social, and environmental dimensions [30]. In this context, ESG criteria have emerged as a fundamental reference framework for evaluating and measuring SC, providing quantifiable indicators aligned with the expectations of investors, regulators, and consumers [31,32].

The transition toward ESG emphasizes a comprehensive sustainability approach that integrates environmental, social, and governance practices [33]. According to Hanson et al. [34], environmental factors assess how a company manages its impact and contributes to environmental protection. The social pillar addresses the management of relationships with stakeholders, such as employees, suppliers, customers, and local communities. Governance factors focus on corporate management, considering shareholder interests and the effectiveness of internal controls. Furthermore, the ESG approach evolves from CSR by adopting a more dynamic perspective that considers both internal and external factors of the company [35].

In sectors such as tourism, it is essential to analyze how this transition is reflected in practical strategies that align sustainable objectives with stakeholder expectations and business competitiveness. The adoption of ESG criteria provides tangible value to companies, as it not only strengthens financial stability but also enhances profitability, risk management, and market value and promotes sustainable practices [36–38]. Tourism companies face specific challenges in this transition [7] due to their high environmental impact and dependence on natural resources, which necessitates more robust and measurable sustainability strategies [39–41].

One of the key approaches for strengthening ESG integration in tourism is the adoption of circular economy (CE) principles. This perspective, aligned with the environmental pillar of ESG, has gained relevance by promoting sustainable processes that minimize the use of non-renewable resources and transform waste into recyclable materials that can be reintegrated into the system [42]. According to Strippoli et al. [43], the implementation of CE strategies in tourism includes plans to "reduce" (eliminating unnecessary plastics), "recycle" (reusing PET), and "reuse" (using refillable products) when conducting plastic waste management. For food waste, key strategies include redistributing unserved food, converting organic waste into compost, and generating biogas. In water management, practices such as greywater recycling, rainwater harvesting, and reducing leaks in distribution systems are essential. These strategies not only foster sustainability but also contribute

directly to the SDGs, particularly responsible consumption (SDG 12), clean water (SDG 6), and life below water (SDG 14).

On the other hand, Gül et al. [44] highlight that the adoption of circular economy principles in tourism, such as replacing primary materials with recycled ones, can significantly reduce fossil fuel and resource consumption, particularly in regards to furniture, hygiene products, and packaging. Moreover, tourism is particularly vulnerable to socioeconomic changes, such as economic crises or health emergencies, including the COVID-19 pandemic, due to its reliance on travelers' discretionary spending [45]. Consequently, the integration of responsible and sustainable practices, such as those promoted by the circular economy, is essential to ensuring the sector's resilience and its alignment with global sustainability goals.

Despite advances in the literature regarding CSR, CS, and ESG in tourism, research gaps persist, hindering a comprehensive understanding of the field. Most studies have a limited scope, focusing on specific subsectors, destinations, or isolated sustainability issues [46]. In this context, a computational literature review, such as the one presented in this study, is essential for synthesizing and integrating existing knowledge, providing a clearer perspective on the development and current state of the field [47–49].

This approach allows for validating whether or not there is a shift in the research focus from CSR to CS, materialized through ESG criteria, given that research on ESG remains insufficient, and academic and business efforts have yet to complete the transition from CSR to a comprehensive ESG framework [8]. According to Miller and Twining-Ward [15], a solid theoretical understanding can foster consistent changes in academia and business practice, reducing ambiguity in the implementation of sustainable strategies. Therefore, by providing an updated and detailed perspective on theoretical dynamics in CSR and CS, this study not only contributes to academic knowledge but also guides the adoption of more effective and sustainable business practices in the tourism sector.

2.2. Application of Data Science in CSR and CS Reports

Corporate reports represent the most effective channel for communicating companies' economic, social, and environmental performance. Over recent decades, various types of corporate reports have emerged, including annual or management reports; integrated reports; and sustainability, CSR, or corporate governance reports [50]. These documents allow companies to declare their commitment to sustainable development while communicating actions and strategies related to CSR and CS [51,52]. The disclosure of such topics is a key tool for mitigating risks in the tourism sector, as it not only reduces economic, political, and climate uncertainties but also serves as an indicator of sustainability and business resilience in a dynamic environment [53]. According to Font et al. [2], the quantity and quality of CSR reports are essential indicators of the degree of responsibility of an industry.

However, the analysis of CSR and CS reports in tourism has traditionally relied on manual and qualitative approaches, limiting the ability to identify complex patterns and conduct large-scale analyses. For instance, Johann [54] conducted a documentary analysis to examine sustainability reports from the TUI group, evaluating corporate strategies implemented during the pandemic. Zientara et al. [55] analyzed 81 CSR reports regarding tourism and hospitality through manual coding, focusing on key aspects such as freedom of association and collective bargaining.

In specific subsectors, similar results are observed. In the hospitality industry, Gajjar and Okumus [56] applied a content analysis approach to evaluate diversity policies in leading companies, identifying practices such as intergenerational benefits and training programs. Quintás et al. [57] analyzed sustainability reports from 54 Spanish hotel chains using

content analysis and the environmental business models (EBM) framework, examining key dimensions such as the creation and capture of environmental value.

In the cruise subsector, Geerts and Dooms [58] conducted a qualitative analysis of sustainability reports, classifying initiatives based on their alignment with core or peripheral activities and primary or secondary stakeholders. Di Vaio et al. [59] reviewed sustainability reports and corporate websites of cruise companies through manual content analysis, identifying 105 practices related to the SDGs, particularly SDG 17. Another study by Di Vaio et al. [60] utilized software like Leximancer (ver. 5.0) to analyze sustainability reports, focusing on specific practices such as plastic reduction and the use of wastewater purification technologies.

While these studies provide valuable insights, their reliance on manual methodologies limits their replicability and complicates large-scale comparisons. However, some recent work has begun to adopt advanced techniques for analyzing tourism reports, representing a step toward more systematic and scalable approaches. For instance, Uyar et al. [12] employed text mining and network analysis to identify thematic changes in sustainability reports between 1999 and 2018, highlighting a transition from a social focus to a balance between social and environmental topics. Similarly, Yoon et al. [61] used techniques such as TF-IDF, topic modeling (LDA), and semantic networks to analyze sustainability reports from Korean companies (2019–2022), identifying thematic trends aligned with the SDGs. These methodologies enable the analysis of large volumes of data, yielding more detailed and precise results. Nonetheless, research in this area remains limited, underscoring the need for further exploration in this field.

3. Research Questions and Hypotheses

This study aims to analyze the perception and evolution of CSR and CS in tourism. To achieve this, the study is structured into two parts: a theoretical analysis and an empirical analysis, each guided by specific questions and hypotheses that address both conceptual dynamics and business practices.

In the first part, a theoretical analysis is conducted through a computational literature review that employs topic modeling to identify key patterns and evaluate their evolution over time. This approach explores whether a transition from CSR to CS has occurred in the academic domain. The associated research questions are as follows:

RQ1. *What topics have been studied in the existing literature on corporate social responsibility and corporate sustainability in tourism over the past twenty years?*

RQ2. *Is there a theoretical paradigm shift from corporate social responsibility to corporate sustainability in tourism?*

These exploratory questions aim to map the most relevant research areas and identify potential changes in research trends. Their answers provide scholars and policymakers with a framework to understand how the theoretical focus on tourism sustainability has evolved.

In the second part, an empirical analysis of 364 corporate reports from Spanish tourism companies is conducted to explore the perception and evolution of CSR and CS practices in this sector. This analysis combines topic modeling using BERTopic, text mining to evaluate the relative frequency of key terms, and the FinBERT model to classify the content into nine ESG categories. The questions guiding this analysis are as follows:

RQ3. *What are the key areas related to the management of corporate social responsibility and corporate sustainability in the corporate reports of tourism companies?*

RQ4. *What importance do tourism companies and subsectors attribute to ESG criteria in the corporate reports of the Spanish tourism sector?*

These questions aim to identify business strategies and priorities in sustainability, highlighting differences between subsectors and contributing to a more detailed understanding of practical trends in the sector.

Additionally, two research hypotheses are formulated to complement the questions posed, enabling a quantitative analysis to evaluate and validate the existence of significant changes in CSR and CS categories over time. In this context, a "significant change" is defined as a statistically relevant variation in the relative frequency of terms associated with each category, evaluated using non-parametric techniques such as bootstrapping and effect size calculation. The associated research hypotheses are as follows:

H1. *There is a significant change over time in the corporate social responsibility category observed through the terms used in the corporate reports of tourism companies.*

H2. *There is a significant change over time in the corporate sustainability category observed through the terms used in the corporate reports of tourism companies.*

These hypotheses aim not only to validate whether or not the observed changes are statistically significant but also to contribute to a quantitative evaluation of the shift in focus between CSR and CS in Spanish tourism companies.

The questions and hypotheses serve complementary purposes. The exploratory questions guide the descriptive analysis and identify general patterns in the literature and reports. On the other hand, the hypotheses take a confirmatory approach, allowing for a quantitative assessment of whether the observed changes in CSR and CS categories are statistically significant. This integrated approach provides a robust framework for evaluating the evolution of CS in tourism.

4. Research Methodology

The methodology of this study combines advanced computational techniques and statistical approaches to analyze the evolution of CSR and CS in the tourism sector. It is structured into two parts: a computational literature review (CLR), which analyzes theoretical patterns and trends, and an empirical study, which examines corporate reports from Spanish tourism companies.

4.1. Methodology Applicable to the Theoretical Approach of the Computational Literature Review

The CLR employs text mining techniques and algorithms to analyze and synthesize large volumes of scientific literature in a specific field [62]. The phases of this methodology are as follows:

4.1.1. Establishing the Conceptual Objective

The purpose of the CLR is to address research questions RQ1 and RQ2, focusing on the predominant topics in the tourism literature regarding CSR and CS, as well as the potential transition from a theoretical paradigm of CSR to CS. This longitudinal analysis allows for identifying how academic attention has evolved in these areas.

4.1.2. Review Planning

The bibliographic search was conducted using the Scopus database, recognized for its high-quality standards [63]. A temporal analysis framework from 2004 to 2023 was established, as publications prior to this period are limited and could introduce biases in the analysis. The search was conducted in English, using key concepts and quotation marks to limit results to those that include the exact expressions selected. The logical operator OR was used to include variations of the concepts, while AND was used to

combine different areas of interest. The search was configured as follows: (("corporate social responsibility" OR "CSR") OR ("corporate sustainability" OR "holistic sustainability" OR "ESG" OR "environmental, social and governance" OR "sustainable strategy" OR "Sustainable Business")) AND ("tourism" OR "travel" OR "hospitality"). This search yielded 1505 documents, including articles, conference proceedings, books, and book chapters, ensuring adequate representation of the field of study.

4.1.3. Selection of Computational Techniques

Topic modeling was selected as the primary technique, as it is a valuable tool for literature reviews in fields with a high volume of publications, enabling the automated and precise identification of key topics based on abstracts [63]. Specifically, the BERTopic model was chosen due to its ability to capture the semantic context of texts [64] and its superior topic coherence and adaptability compared to those of classical approaches. For instance, latent semantic analysis (LSA) [65] relies on linear techniques that struggle to capture complex semantic nuances and require intensive preprocessing. Latent Dirichlet allocation (LDA) [66], while widely used, assumes a "bag-of-words" representation that disregards semantic relationships between terms, limiting topic differentiation. The bigram topic model [67] improves upon this by incorporating dependencies between consecutive words, but its high computational cost makes it less viable for large datasets.

BERTopic overcomes these limitations by separating the embedding generation from the topic modeling process, integrating more advanced language models and adapting to future developments in natural language processing [68]. However, its keyword-based approach may be less interpretable in certain contexts, emphasizing the need for manual validation to ensure the relevance and accuracy of the results.

This model enabled the analysis of thematic evolution over time, aligning directly with the study's objectives.

4.1.4. Data Processing

The preprocessing of textual data was performed using the Python programming language (ver. 3.10.9), applying advanced natural language processing (NLP) techniques. This step is essential for structuring and normalizing the data to ensure that the analysis algorithms operate accurately and consistently. Following the methods of Uysal and Gunal [69], the following stages were implemented:

- Tokenization: The text was divided into meaningful units, called tokens, which can be words or phrases. During this process, non-informative elements, such as punctuation marks, special characters, and whitespace, were removed to focus the analysis on relevant terms.
- Stop word removal: Common grammatical terms, such as articles, prepositions, and conjunctions, were excluded, as they do not contribute analytical value in the context of topic modeling and text classification.
- Lowercasing: All terms were converted to lowercase to ensure uniform representation and avoid inconsistencies caused by capitalization differences.
- Lemmatization: Words were reduced to their base or root form, unifying morphological variants of the same concept and minimizing the dispersion of textual data.

4.1.5. Content Analysis

Content analysis was conducted using the BERTopic machine learning technique, which organizes texts into key topics based on their semantic similarity. To identify the most representative words for each topic, this model employs the c-TF–IDF metric (class-based term frequency–inverse document frequency), a tool that evaluates the relevance

of terms within each thematic group relative to the entire corpus. This metric highlights the distinctive terms of each topic, ensuring they reflect consistent patterns in academic discussions. Applying this model allowed for the identification of key dynamics and recurring areas in academic literature over the past two decades, providing a comprehensive view of the field of study. Notably, the empirical methodology section delves into additional aspects of the BERTopic model, including comparisons with other models, which supports and justifies its selection as the analysis tool for this research.

4.1.6. Validation and Synthesis of Results

The identified topics were validated through a manual categorization process to group them into broader research areas. Following recent approaches in topic modeling studies using BERTopic [70–72], a qualitative evaluation was conducted to verify the semantic coherence of the topics. Specifically, the keywords generated by BERTopic were reviewed to determine whether they represented a homogeneous conceptual framework, and representative publications for each topic were manually inspected to confirm their alignment with the assigned category.

Additionally, a descriptive analysis based on relative proportions was performed, understood as the percentage of publications associated with CSR and SC topics relative to the total number of publications analyzed each year. This approach, complemented by the non-parametric Mann–Kendall test [73,74], allowed for the detection of monotonic trends, meaning sustained patterns of increase or decrease in the data, without assuming a linear relationship. This method is particularly appropriate because it does not require the data to follow a normal distribution, which is essential when working with short time series or relative proportions. Furthermore, the test is robust against outliers, ensuring that results are not disproportionately affected by extreme data points. In this way, clear insights were obtained regarding the evolution of CSR and SC themes in academic literature, aligning with the study's objectives.

4.2. Methodology Applicable to the Empirical Approach of the Review of Corporate Reports from Spanish Tourism Companies

The empirical research is based on the analysis of annual reports, CSR and sustainability reports, and integrated reports from Spanish tourism companies. These strategic documents contain key information about corporate practices related to CSR, CS, and ESG [75]. To interpret and analyze these reports, content analysis is employed—a textual methodology that identifies patterns and meanings from structured texts [76]. This approach has been widely used in studies on sustainability and CSR [77–82] and is highlighted in this research for its automated implementation using data science techniques.

This analysis aligns with research questions RQ3 and RQ4, which aim to identify key management areas in CSR and CS, as well as the relevance of ESG criteria in corporate reports across subsectors (transport, hotels and accommodations, and restaurants). It also addresses hypotheses H1 and H2, statistically evaluating whether there has been a significant change in the relative frequencies of terms associated with CSR and CS over time. As in the theoretical study, the preprocessing, analysis, and evaluation of the information are conducted using Python, enabling the efficient structuring and analysis of large volumes of textual data. Figure 1 presents the methodological framework of the empirical study.
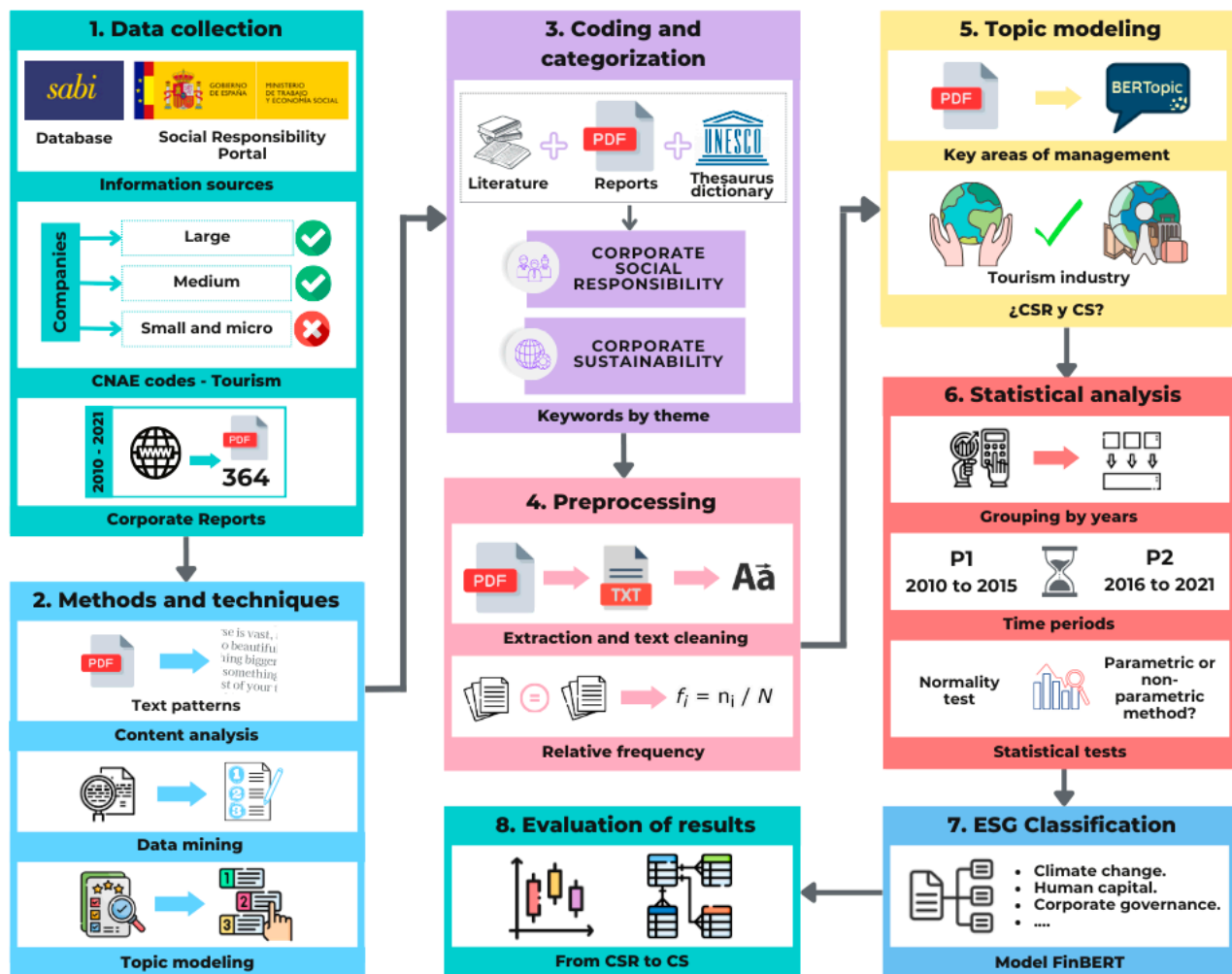
Figure 1. Methodological framework of empirical research.

4.2.1. Data Collection

The selection of companies was conducted in April 2022 using the Iberian Balance Sheet Analysis System [83] database. The CNAE codes associated with tourism activities (hotels and accommodations, transportation, and restaurants) were filtered, as were active companies with a functioning website. Subsequently, the companies were classified into three groups based on their operating revenue for the most recent year available: large (more than EUR 50 million), medium (between EUR 10 million and EUR 50 million), and small or micro (less than EUR 10 million).

For this study, only large and medium-sized companies (435) were selected, as these have a higher likelihood of publishing corporate reports due to their visibility and the pressure from stakeholders to disclose their economic, social, and environmental performance [84]. Previous studies have demonstrated a positive relationship between organizational size and the quality of CSR and sustainability disclosures, confirming that larger companies tend to provide more detailed reports [84–86].

Additionally, these companies are subject to stricter regulatory obligations. For instance, the Corporate Sustainability Reporting Directive (CSRD), adopted in 2022, requires large and some medium-sized companies to disclose ESG-related information using the European Sustainability Reporting Standards (ESRS). This directive, which replaces the previous Non-Financial Reporting Directive (NFRD), elevates sustainability to the level of financial information, promoting greater transparency and comparability. Similarly, Spain's Climate Change and Energy Transition Law 7/2021 mandates companies to identify climate

risks associated with their activities, set specific decarbonization goals, and calculate—and optionally offset—their emissions. These regulations complement the objectives of the European Green Deal and the "Fit for 55" legislative package, which aims to reduce emissions by 55% by 2030.

The search for reports was supplemented with information from Spain's Social Responsibility Portal [87]. A total of 404 reports published between 2000 and 2021 were downloaded. However, data from 2000 to 2009 were insufficient for applying relevant statistical analyses and comparisons. For this reason, the analysis period was set from 2010 to 2021, comprising 364 corporate reports divided into two periods: P1 (2010–2015) and P2 (2016–2021). This division allows for comparisons across equivalent durations and is justified by the adoption of the Sustainable Development Goals (SDGs) in 2015, marking a significant shift in corporate policies and strategies toward sustainability [88]. The tourism sector, in particular, has been identified as one of the most engaged in regards to this initiative [89].

4.2.2. Methods and Techniques of Investigation

This study employed three complementary text analysis techniques: word frequency, topic modeling, and classification using FinBERT. Each technique was strategically selected for its ability to address different aspects of the research objectives.

Word frequency was chosen as the primary text mining method due to its ability to identify linguistic usage patterns and quantify the prominence of specific concepts in large text corpora [90]. This approach enabled the establishment of direct comparative metrics between CSR and CS themes, considering that the relative frequency of associated terms reflects their thematic relevance in the texts. While alternative methods, such as collocation analysis or the use of n-grams, could capture more complex relationships, word frequency was selected for its simplicity, interpretability, and ability to provide a clear initial overview of the thematic distribution.

As a complementary analysis, topic modeling was used to uncover latent thematic structures in the texts. This technique identifies co-occurrences and more complex semantic associations not captured by word frequency. BERTopic was chosen for its ability to model semantic context and generate coherent thematic groupings using the c-TF–IDF metric, designed to highlight representative terms in each thematic cluster [64]. Compared to alternative techniques such as latent semantic analysis (LSA) or word embeddings, BERTopic stands out for its interpretability and adaptability to longitudinal analysis, making it particularly suitable for tracking thematic evolution in reports.

Finally, FinBERT, a language model specifically trained on financial texts [91], was used to classify the texts into nine ESG categories. Its selection over generalist models like BERT base or RoBERTa was justified by its high accuracy and pretraining in financial contexts, ensuring superior performance in classifying corporate reports. This capability is essential for capturing the specific nuances of financial language used in the tourism reports analyzed.

4.2.3. Coding and Categorization

In the third phase, representative keywords for CSR and CS were defined through a structured coding and categorization process. This process was based on a deductive–inductive approach [92] and considered three main elements:

- Theoretical grounding in previous studies: The process began with relevant research such as that of Karen [93], Palazzo and Scherer [94], and Van Marrewijk [20], which identify recurring terms in the literature regarding CSR and CS. These references

provided a robust initial framework for identifying keywords, particularly in an interdisciplinary context like tourism.

- Supplementation with the UNESCO Thesaurus: Concepts related to "social responsibility" and "sustainable development" from the UNESCO Thesaurus [95] were incorporated. This thesaurus offers a controlled and structured list of terms used in international research, ensuring broad and up-to-date coverage of concepts, while minimizing biases associated with the use of individual sources.
- Contextual validation for Spanish companies: To adapt the keywords to the Spanish business context, a qualitative analysis of the collected reports was conducted, identifying specific terms used by companies in their corporate reports. This analysis enriched the initial categorization with terminology specific to the local context, resulting in a categorized framework that integrates both international academic perspectives and Spanish business practices.

Figure 2 presents the resulting categorization, including the identified terms for both categories.
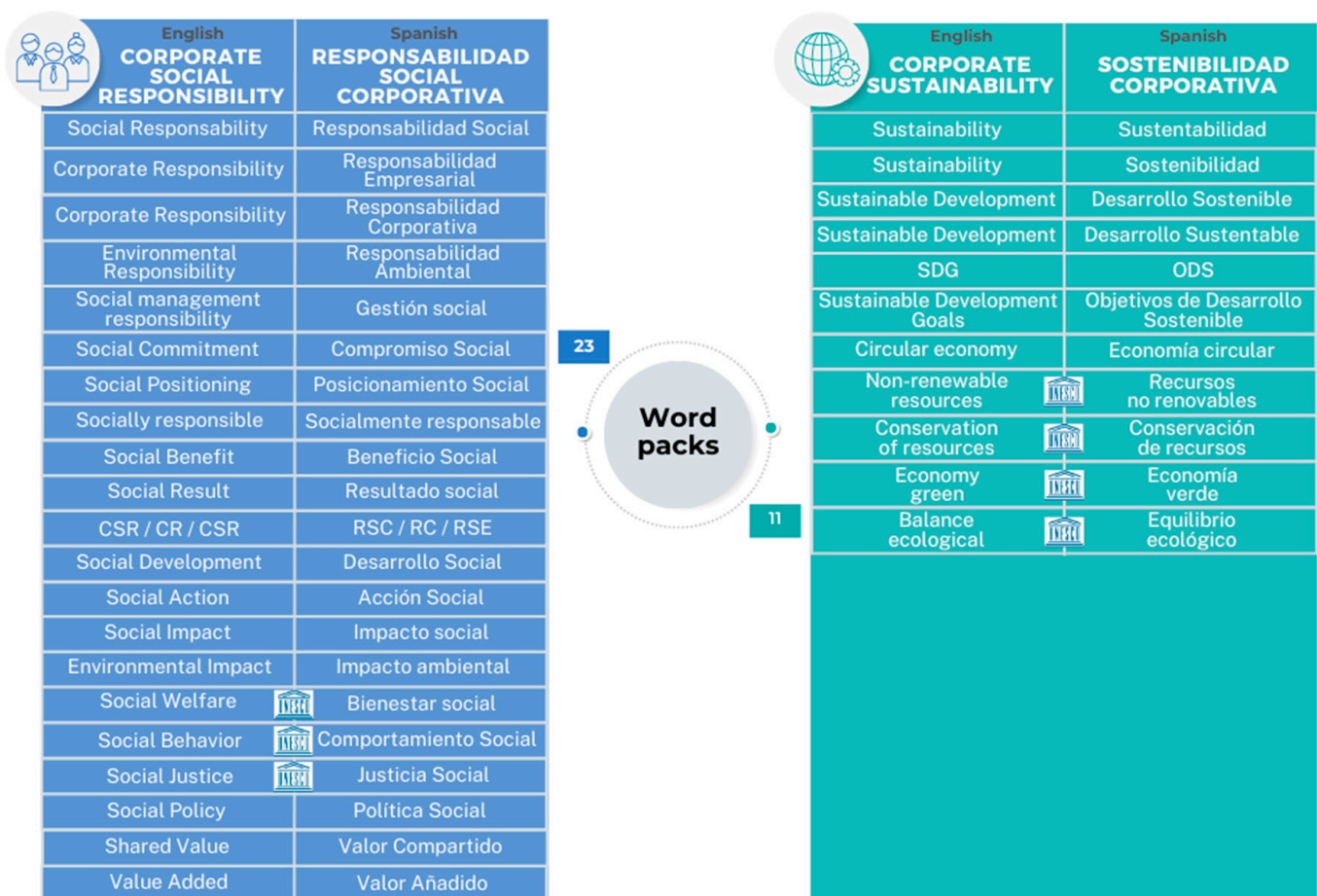

Figure 2. Categorization of CSR and CS.

4.2.4. Preprocessing

The preprocessing of corporate reports was a crucial step to ensure the quality and reliability of subsequent analyses. This process is essential for text extraction algorithms to function correctly, minimizing errors and maximizing the validity of the results. According to García et al. [96], data preparation includes techniques such as transformation, normalization, integration, and cleaning, which were adapted in this case to address the variations in format, structure, and length characteristics of the corporate documents analyzed.

Following a procedure similar to that used in the theoretical approach of the computational literature review, several specific steps were carried out:

- Conversion and redundancy removal: PDF files were converted to the TXT format, and non-analytical elements such as headers, footers, and watermarks were removed.
- Text cleaning and normalization: Non-informative patterns, such as symbols, numbers, and isolated characters, were removed, and all text was converted to lowercase to unify the representation of terms and avoid inconsistencies.
- Stop word removal and lemmatization: Stop words, such as articles, prepositions, and conjunctions, which do not contribute analytical value, were eliminated. Additionally, lemmatization was performed to reduce words to their base or root form, unifying morphological variants and improving the semantic coherence of the data.
- Adjustment using relative frequency: To ensure comparability between documents of different lengths, the relative frequency method [97] was applied. Absolute word frequencies were adjusted to the total size of the text by dividing the absolute frequency of each pattern or word by the total number of words in the document, as proposed by Salinas and Martínez [98]. This approach has proven effective in previous research [76], particularly when document sizes vary considerably.

These NLP tasks ensured that the textual data were standardized and prepared for subsequent semantic analyses, reducing biases stemming from the heterogeneity in the formats and styles of corporate reports.

4.2.5. Topic Modeling Based on the BERTopic Approach

At this stage, BERTopic [64] was used to identify key areas in the management of CSR and SC. This procedure follows the approach previously described in Section 4.1.3, where the selection of BERTopic as the topic modeling method and its comparison with other approaches, such as LDA, are detailed. As mentioned, BERTopic has demonstrated superior performance in topic identification by leveraging language embeddings and dynamic clustering, making it particularly suitable for this analysis.

To ensure that the results were consistent with the objectives of the empirical study, a guided approach was adopted by configuring seed topics related to CSR and SC. While this method enhances the relevance and interpretability of the results, it could potentially introduce bias by prioritizing certain topics. To mitigate this risk, a comprehensive manual review of the identified topics was conducted, verifying that they accurately reflected the content of the reports and did not exclude relevant areas of interest.

4.2.6. Statistical Analysis of the Data

In this phase, the relative frequencies of words associated with CSR and CS themes were calculated by dividing the absolute frequency of each term by the total number of words in each document. This approach adjusted the metrics to the text size, reducing potential biases arising from variability in the length of the reports. Subsequently, these relative frequencies were consolidated by year and category, creating a longitudinal database for statistical analysis.

To identify significant differences between categories and periods, data normality was assessed using the D'Agostino test [99], which tested the null hypothesis that the data followed a normal distribution. The results indicated that the data did not meet this assumption (p-value < 0.05), justifying the use of non-parametric methods. Bootstrapping [100] was employed as the primary technique due to its ability to handle non-normal data. This non-parametric approach uses random resampling with replacement to generate sample distributions, enabling the calculation of robust confidence intervals and p-values. In this study, 9999 resampling iterations were performed to ensure the precision and stability of

the estimates. Additionally, effect sizes, such as Cohen's d, were calculated to complement the statistical evaluation with a measure of the magnitude of the observed changes.

This comprehensive statistical approach provided a solid foundation for evaluating the proposed hypotheses, offering a detailed understanding of the differences and trends in CSR and CS categories over time.

4.2.7. Classification Under ESG Criteria

To detail the sustainable practices communicated by tourism companies in their corporate reports over time, FinBERT, a pre-trained language model specifically designed for the financial domain [91], was applied. FinBERT, based on the BERT architecture, has been pre-trained on financial texts, including annual and quarterly corporate reports, financial analyst reports, and earnings call transcripts. This model has demonstrated outstanding performance in ESG topic classification, achieving an accuracy of 89.5%, outperforming other machine learning algorithms such as naive Bayes (80.0%), support vector machine (SVM) (75.0%), convolutional neural network (CNN) (82.0%), and slightly surpassing the performance of BERT (87.0%).

Thus, FinBERT is considered an ideal model for ESG criteria classification in this study due to its specialized training on financial texts, its robustness, and its high performance. Moreover, FinBERT has been successfully used in previous studies [101–109]. In this study, the pre-trained version available in the Hugging Face library was used without additional fine-tuning, given its validated performance in the financial domain. This enabled the classification of corporate report texts into nine ESG categories: climate change, natural capital, pollution and waste, human capital, product responsibility, community relations, corporate governance, ethics and corporate values, and non-ESG. Texts were segmented into sentences and classified into these categories, generating relative scores that reflected the proportion of content assigned to each.

Although no specific adjustments were made for the Spanish tourism context, this approach enabled an efficient analysis of sustainable practices by leveraging the robustness and flexibility of the pretrained model. Since FinBERT has been previously trained on large volumes of financial texts and validated in similar studies, no additional accuracy tests were conducted for this dataset. Instead, a manual review of a sample of the generated classifications was performed to ensure that the assigned categories corresponded to the text content and coherently reflected ESG themes. While this direct implementation exhibits limitations in capturing nuances specific to the Spanish context, it provides a solid foundation for analysis and represents an opportunity for future studies aiming to adapt the model to the specific needs of the Spanish tourism sector.

5. Results

5.1. Results of the Theoretical Approach: Computational Literature Review

Topic modeling in the computational literature review identified 20 themes that comprise the main research areas related to CSR and CS in tourism (Figure 3). The primary theme identified, tourism development (T0), highlights the importance of rural and community tourism as strategies to promote sustainability in the tourism sector. Key terms such as community, rural, sustainable, and destination reflect a focus on initiatives that prioritize local community engagement and responsible destination management. According to Lane and Kastenholz [110], rural tourism is a tool for sustainable development that requires a broader and more collaborative approach among researchers, entrepreneurs, and local communities.
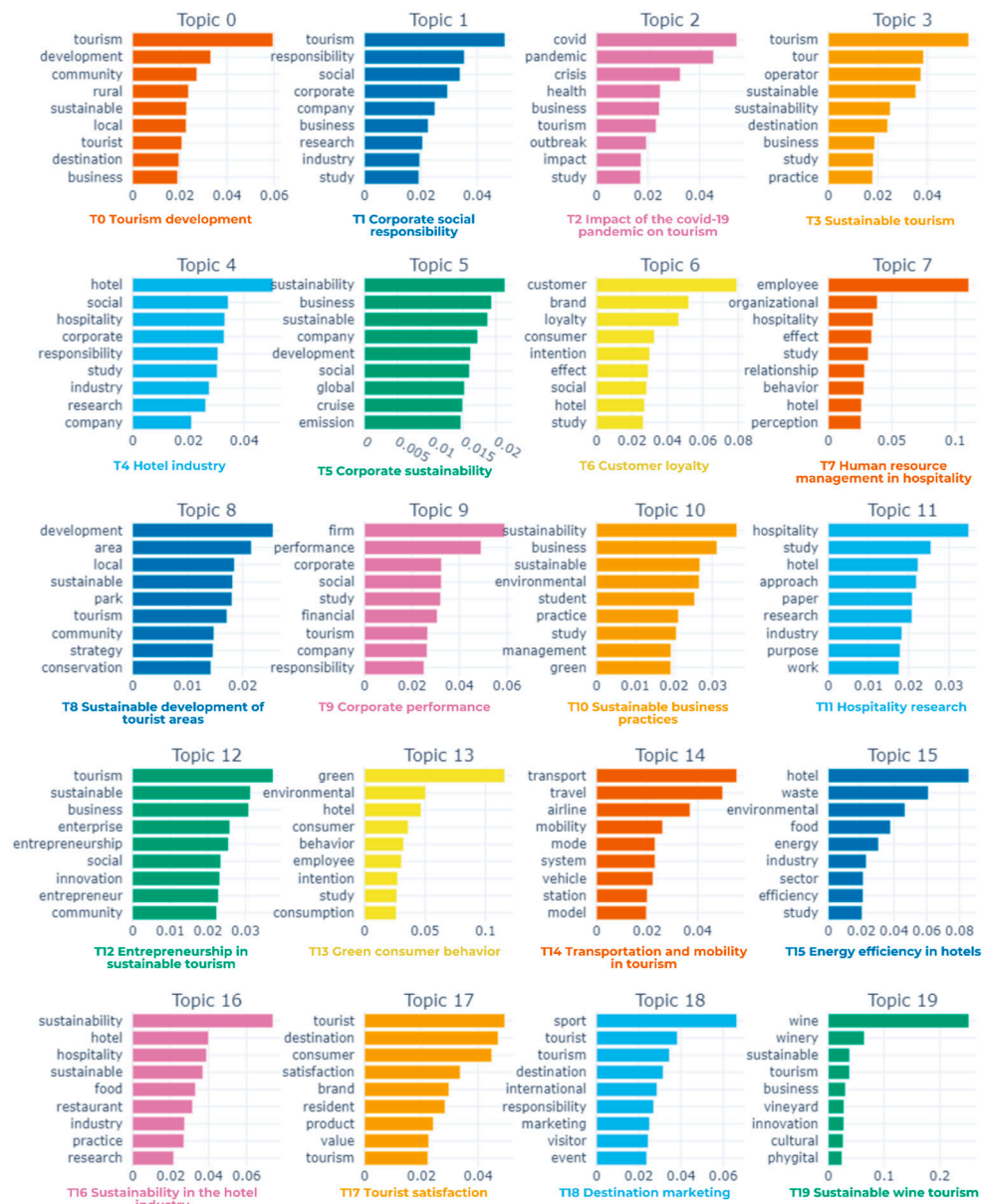
Figure 3. Research topics associated with CSR and CS in tourism.

The theme corporate social responsibility (T1) reflects ongoing interest in ethical and responsible business practices aimed at strengthening corporate legitimacy and responding to stakeholder demands [3]. The theme impact of the COVID-19 pandemic (T2) underscores how terms such as pandemic, crisis, health, and business have dominated recent research, reflecting a more sustainable, ethical, and resilient focus within the sector [111].

Topics such as sustainable tourism (T3) and corporate sustainability (T5) emphasize the importance of integrated management in tourism, as illustrated by key terms such as sustainability, destination, global, and emission. These terms underscore the need to simultaneously address the economic, social, and environmental dimensions of sustainability [112]. Furthermore, CS can act as a catalyst for community development and cooperation among key stakeholders, fostering relationships built on trust and encouraging responsible practices [113].

Other notable topics, such as the hotel industry (T4) and human resource management in hospitality (T7), emphasize the need to integrate sustainability and social responsibility into operational management. Sakshi et al. [114] point out that environmental policies and

sustainability practices in the hotel industry have a significant and positive impact on both the environmental and financial performance of hotels, demonstrating the feasibility of these practices for the subsector. Similarly, customer loyalty (T6), represented by terms such as customer, loyalty, and brand, highlights a focus on sustainable strategies aimed at improving consumer satisfaction and retention. Han et al. [115] show that practices such as water conservation and energy efficiency, combined with positive guest experiences, enhance guest loyalty and ensure the long-term success of sustainable hotels.

Finally, emerging topics such as sustainable tourism entrepreneurship (T12) and sustainable wine tourism (T19), represented by key terms like entrepreneurship, sustainability, social, innovation, community, wine, and business, underscore the role of innovation and business practices in developing sustainable tourism niches. De Lange and Dodds [116] emphasize that social entrepreneurship drives sustainable tourism by fostering sustainable enterprises, increasing competitiveness, facilitating environmental and social regulations, and promoting local economic development with an international outlook. Dyah Utami et al. [117] highlight that rural tourism entrepreneurship, based on collaboration, innovation, and sustainable resource management, is essential for developing sustainable tourist villages. In the case of wine tourism, Vecchio et al. [118] highlight the growing interest among young adults in sustainable winery tourism, driven by factors such as environmental attitudes, social norms, and engagement with wine.

Table 1 synthesizes and consolidates the identified topics into eight major research categories. The core themes of this research (CSR and CS) are established as independent categories, while other strong areas are highlighted, such as "sustainable tourism and development", "hospitality and management", "customers and marketing", "corporate performance", "COVID-19 pandemic", and "transport and mobility".

Table 1. Thematic categories of research on CSR and CS in tourism.

Category		Topic	
C0	Corporate Social Responsibility	T1	Corporate social responsibility
C1	Corporate Sustainability	T5	Corporate sustainability
		T10	Sustainable business practices
		T0	Tourism development
		T3	Sustainable tourism
C2	Sustainable Tourism and Development	T8	Sustainable development of tourist areas
		T12	Entrepreneurship in sustainable tourism
		T19	Sustainable wine tourism
		T4	Hotel industry
		T7	Human resource management in hospitality
C3	Hospitality and Hotel Management	T11	Hospitality research
		T16	Sustainability in the hotel industry
		T15	Energy efficiency in hotels
		T6	Customer loyalty
C4	Customers and Marketing	T17	Tourist satisfaction
		T13	Green consumer behavior
		T18	Destination marketing
C5	Corporate Performance	T9	Corporate performance
C6	COVID-19 Pandemic	T2	Impact of the COVID-19 pandemic on tourism
C7	Transportation and Mobility	T14	Transportation and mobility in tourism

To better understand the structure and evolution of tourism research regarding CSR and CS, Figure 4 presents the number of publications per topic over time. This visualization highlights the evolution of interest in specific topics and their annual concentration in scientific output. An increase in the number of publications is evident across most topics

starting in 2018, peaking in the most recent years (2021–2023), reflecting the growing focus of tourism literature on sustainability and social responsibility during this period [21,22].
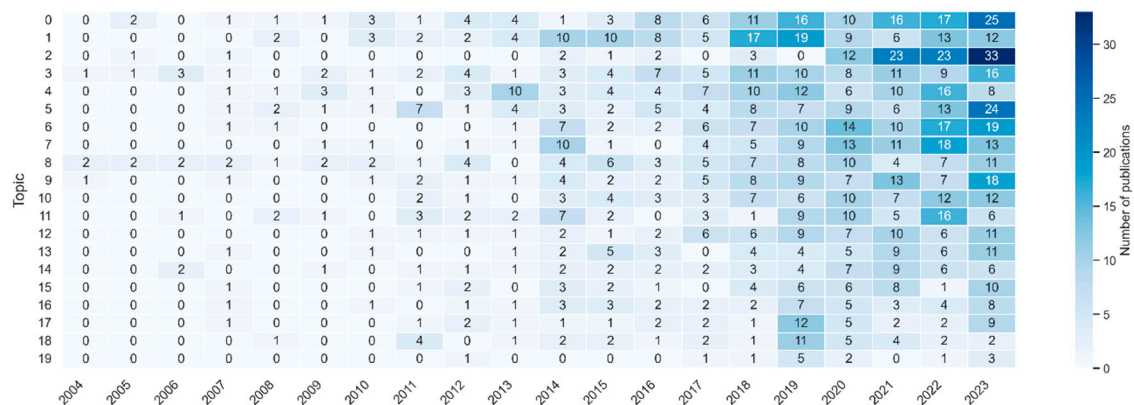
Topic	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
0	0	2	0	1	1	1	3	1	4	4	1	3	8	6	11	16	10	16	17	25
1	0	0	0	0	2	0	3	2	2	4	10	10	8	5	17	19	9	6	13	12
2	0	1	0	1	0	0	0	0	0	0	2	1	2	0	3	0	12	23	23	33
3	1	1	3	1	0	2	1	2	4	1	3	4	7	5	11	10	8	11	9	16
4	0	0	0	1	1	3	1	0	3	10	3	4	4	7	10	12	6	10	16	8
5	0	0	0	1	2	1	1	7	1	4	3	2	5	4	8	7	9	6	13	24
6	0	0	0	1	1	0	0	0	0	1	7	2	2	6	7	10	14	10	17	19
7	0	0	0	0	0	1	1	0	1	1	10	1	0	4	5	9	13	11	18	13
8	2	2	2	2	1	2	2	1	4	0	4	6	3	5	7	8	10	4	7	11
9	1	0	0	1	0	0	1	2	1	1	4	2	2	5	8	9	7	13	7	18
10	0	0	0	0	0	0	0	2	1	0	3	4	3	3	7	6	10	7	12	12
11	0	0	1	0	2	1	0	3	2	2	7	2	0	3	1	9	10	5	16	6
12	0	0	0	0	0	0	1	1	1	1	2	1	2	6	6	9	7	10	6	11
13	0	0	0	1	0	0	1	0	0	1	2	5	3	0	4	4	5	9	6	11
14	0	0	2	0	0	1	0	1	1	1	2	2	2	2	3	4	7	9	6	6
15	0	0	0	1	0	0	0	1	2	0	3	2	1	0	4	6	6	8	1	10
16	0	0	0	1	0	0	1	0	1	1	3	3	2	2	2	7	5	3	4	8
17	0	0	0	1	0	0	0	1	2	1	1	1	2	2	1	12	5	2	2	9
18	0	0	0	0	1	0	0	4	0	1	2	2	1	2	1	11	5	4	2	2
19	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	5	2	0	1	3

Figure 4. Evolution of the volume of scientific publications regarding CSR and CS in tourism.

The growth in publications is particularly notable in topics 0 (tourism development), 2 (impact of the COVID-19 pandemic on tourism), and 5 (corporate sustainability), indicating a rising interest in these fields and continuous development in the literature. This trend can be interpreted as a response to the need to adapt tourism research to emerging demands for sustainability and resilience in the face of global challenges such as the COVID-19 pandemic. In contrast, topics 18 (destination marketing) and 19 (sustainable wine tourism) have the lowest number of publications throughout the analyzed period, suggesting that these are emerging areas that remain underexplored or have lesser relevance in the current literature. For instance, Martínez-Falcó et al. [119] highlight that wine tourism acts as a catalyst for improving the sustainable performance of wineries and connects with global trends in sustainability, innovation, and community development. Additionally, tourists are increasingly interested in authentic experiences, such as winery visits, wine tastings, and activities related to local culture.

These findings also align with the study by Huang [120], which emphasizes how sustainability has become a central axis in tourism research. According to this analysis, growing topics such as sustainability, disruptive technologies, and the effects of COVID-19 have gained prominence in recent years, reflecting a shift in sector priorities toward more sustainable and innovative management. Simultaneously, traditional topics like customer satisfaction, service quality, and destination management have evolved into more specific and interdisciplinary approaches, gradually being overshadowed by newer perspectives that respond to the demands of the current tourism context.

To evaluate whether there is a conceptual paradigm shift from CSR to CS in tourism research, a descriptive analysis of the relative proportions of publications was conducted, complemented by the non-parametric Mann–Kendall test [73,74]. Figure 5 presents the relative proportion of publications on CSR and CS between 2014 and 2023, based on the annual total of publications identified for each topic. In the early years of this period, CSR predominated, representing more than 75% of the publications. However, since 2016, a progressive convergence between the two themes has been observed, culminating in the predominance of CS in the last three years. By 2023, CS accounted for 66.7% of the publications, while publications dealing with CSR decreased to 33.3%. This shift suggests a transition toward more comprehensive and sustainable approaches, aligned with the global tourism sector's demands and the Sustainable Development Goals (SDGs).
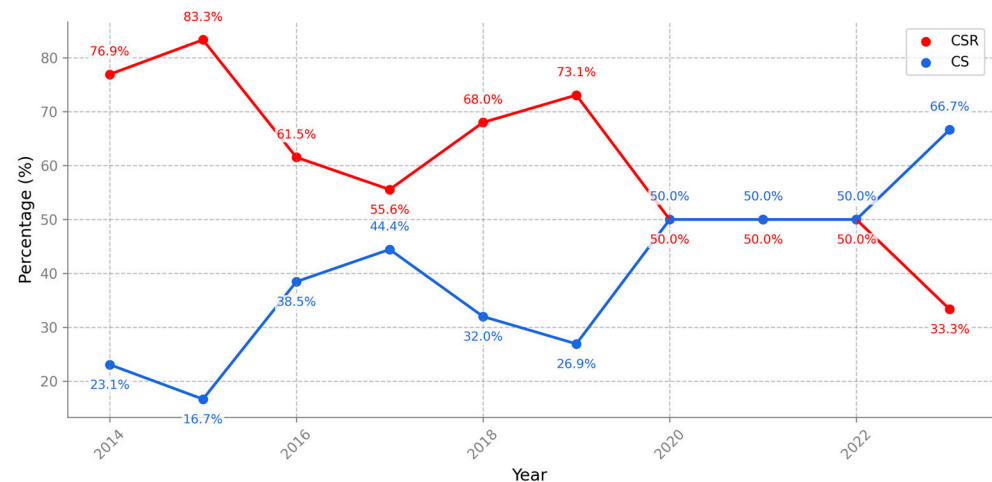
Figure 5. Relative proportion of publications related to CSR and CS (2014–2023).

The Mann–Kendall test confirmed these trends statistically. The results (Table 2) show a statistically significant upward trend in CS (p-value = 0.004; Tau = 0.733), while no significant trend was detected for CSR (p-value = 0.788; Tau = 0.089). The Tau coefficient, which measures the strength and direction of a monotonic trend, is high and positive for CS, indicating a consistent growth pattern in this theme. Conversely, the Tau value for CSR is low, close to zero, suggesting the absence of a significant monotonic trend, reflecting possible stagnation in academic interest in this topic.

Table 2. Mann–Kendall test results for CSR and CS (2014–2023).

Topic	Trend	*p*-Value	Tau
Corporate Social Responsibility	Not significant	0.788	0.089
Corporate Sustainability	Increasing	0.004	0.733

This temporal analysis underscores the growing relevance of CS as an emerging theme in academic literature, particularly in recent years, while CSR shows a more stagnant trend. These results highlight a shift in research priorities, providing a detailed perspective on how academic studies are responding to contemporary sustainability demands in the tourism sector.

5.2. Results of the Empirical Approach: Analysis of Corporate Reports from Tourism Companies

5.2.1. Topic Modeling with the BERTopic Approach

To apply guided topic modeling using BERTopic, three main themes were defined for the model to converge: CSR, CS, and the economic domain. The inclusion of the economic domain reflects the nature of the analyzed data, which includes annual reports and corporate memos where financial results are central. Separating this dimension as an independent topic aimed to mitigate its dominance in the analysis, ensuring that other key aspects related to CS and CSR could be identified. This decision contributed to a more balanced and representative analysis of the report content, comprehensively reflecting the strategic and operational priorities of the companies.

Additionally, the reports were divided into sentences, and specific parameters were configured with a focus on two aspects. First, for dimensionality reduction, UMAP (uniform manifold approximation and projection) was used, configured with 15 neighbors and a projection in five dimensions. These settings were chosen following the methodological recommendations of BERTopic [64] and supported by the study by Sánchez-Franco and

Rey-Moreno [121], which highlights UMAP's ability to preserve both local and global relationships in textual data. The configuration of 15 neighbors ensures the preservation of local structure within the data, while the projection into five dimensions balances simplicity with the retention of relevant information. Additionally, the Euclidean metric was used for its effectiveness in multidimensional analyses, and the parameter random_state = 42 was set to ensure the reproducibility of the results.

Second, for clustering, the k-means method was used, with 50 clusters for each analyzed period (P1 and P2). This value was defined after preliminary iterations that balanced the thematic diversity of the corpus with the clarity of the analysis. This configuration ensured a sufficient number of topics for detailed analysis, while maintaining interpretive coherence and aligning with the study's objectives.

Finally, each of the topics identified by BERTopic was manually titled, based on the most representative words generated by the c-TF–IDF model. While this process included a manual component, a detailed validation was conducted by comparing the proposed titles with the underlying thematic content and verifying their coherence with the priorities identified in the corpus. This approach reduced potential biases and ensured that the titles accurately reflected the study's focus.

The results show that in P1 (2010–2015), CSR ranked seventh and was primarily represented by the words social, responsibility, disability, people, and foundation. Topic 12 was associated with CS and comprised terms such as environmental, sustainability, management, and biodiversity. For P2 (2016–2021), CS was represented in Topic 12, while CSR appeared in Topic 20. Other topics reflected various aspects or areas of company management, such as the subsectors included in the sample (hotels, rail, land, and maritime transport). Additionally, prominent themes emerged related to the economic domain, customers, governance, employees, ethics, inclusion, and equality.

5.2.2. Key Areas of CSR and CS Management

To deepen the results and analyze the study's themes (CSR and CS), a topic similarity matrix was calculated for each period. The matrix values range from 0 to 1, where values closer to 1 indicate high similarity between topics, while those closer to 0 reflect low similarity. To identify significant thematic relationships, a similarity threshold of 0.60 was used, selected based on a statistical analysis of the similarity distributions. This threshold is strategically positioned between the median (0.539 in P1 and 0.578 in P2) and the 75th percentile (0.653 in P1 and 0.668 in P2) of the distribution (see Figure 6). This positioning allows for the identification of robust thematic relationships, while filtering out weaker associations. Its placement, immediately following the peak of the distributions, ensures the exclusion of less relevant connections, while preserving the most significant associations.
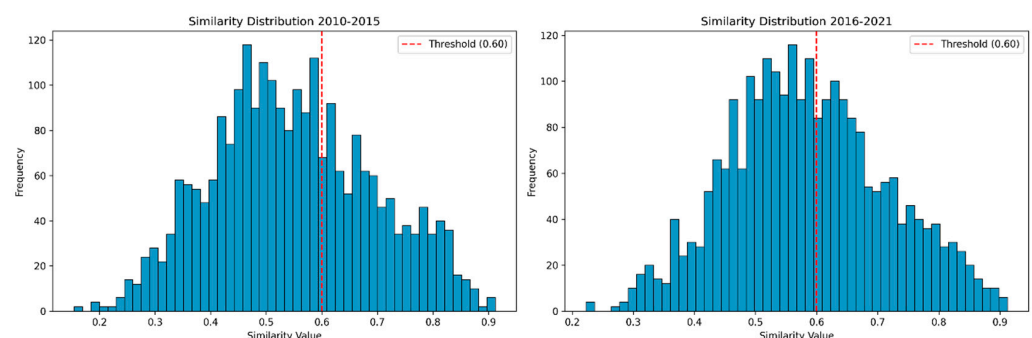

Figure 6. Similarity distributions by periods.

The validity of this threshold is confirmed by its ability to capture the temporal evolution of thematic relationships, while maintaining coherence in topic grouping. In P1,

856 significant connections were identified (34.94% of similarities), whereas in P2, this number increased to 1056 connections (43.10%), reflecting a greater thematic interconnection. A sensitivity analysis with alternative thresholds (0.55 and 0.65) reinforces this choice. With 0.55, the number of connections increases significantly (1162 in P1 and 1442 in P2), but it includes less-relevant associations that could introduce noise. Conversely, with 0.65, the connections decrease (628 in P1 and 724 in P2), potentially excluding important relationships. The 0.60 threshold strikes a balance between topic differentiation and clustering, ensuring that topics remain sufficiently distinct without excessive fragmentation of the thematic space. This guarantees that the identified connections accurately reflect relevant patterns in the evolution of CSR and SC, preserving key information without losing meaningful insights.

The selected topics were classified as related to CSR or CS, based on their predominant similarity. A topic is linked to CSR if its similarity is greater with other CSR topics than with CS topics, and vice versa. This procedure allowed for the identification of specific patterns for each period. Figure 7 presents the similarity matrix for P1, where nine topics were identified as related to CSR and seven topics to CS. In P2, detailed in Figure 8, 12 topics were identified as related to CS and 10 to CSR, reflecting a shift in the areas of emphasis between the two periods.
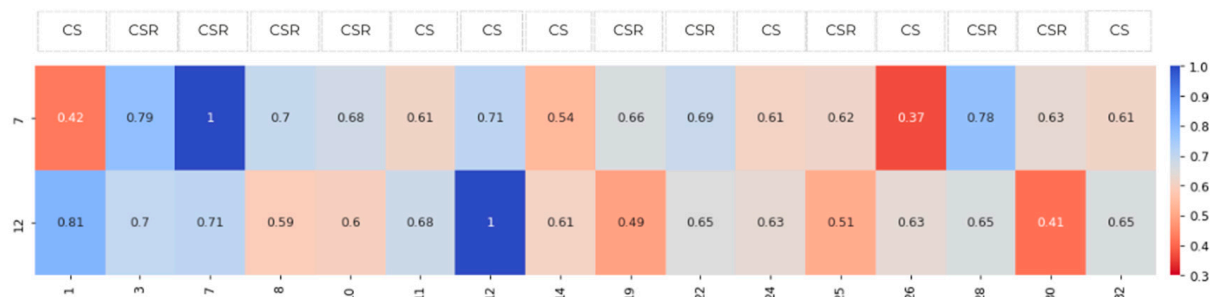
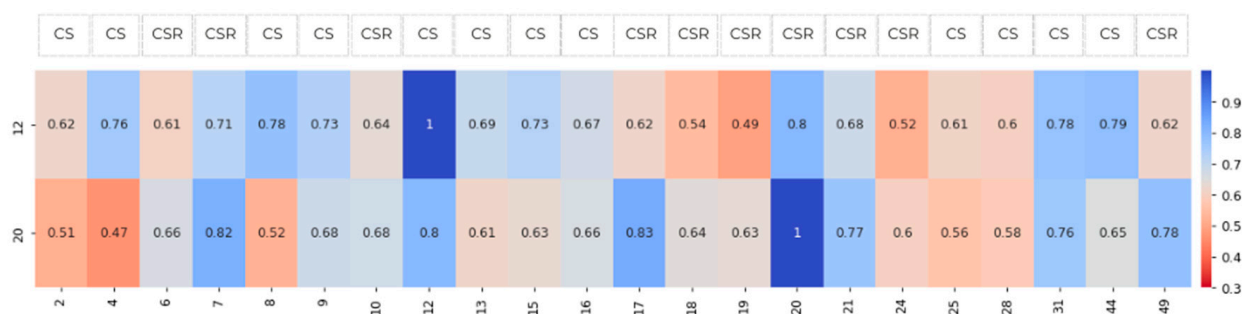**Figure 7.** Topic similarity matrix for CSR and CS in P1.

Figure 8. Topic similarity matrix for CSR and CS in P2.

Figures 9 and 10 present the most representative words from topics linked to CSR in periods P1 and P2, respectively. The results highlight key areas such as organizational management, philanthropy, governance, communication, customers, ethics, and employees, including aspects related to hiring, training, working conditions, equality, diversity, and inclusion. In P1, there is a focus on suppliers as a stakeholder group, while in P2, there is a shift toward social contribution, represented by terms such as community, contributes, and development. These results underscore the importance of strategic management in the tourism sector, which can encompass multiple domains, such as corporate governance, the internal environment of organizations, or strategic decisions [122].

Figure 9. Topics related to CSR in P1.

Figure 10. Topics related to CSR in P2.

The evolution observed between P1 and P2 highlights how management priorities have transitioned from operational approaches toward greater integration of social and community objectives, aligning with contemporary demands for sustainability. Philanthropy, identified as a key area in P1, positions itself as a strategic tool that allows companies to address critical needs in local communities, enhancing their social and political legitimacy

and contributing to sustainable development [123]. This emphasis on social contribution and governance aligns with the CSR approach, which aims to meet ethical and social obligations toward stakeholders by integrating principles of equity and inclusion [5].

On the other hand, Figures 11 and 12 show the topics related to CS in P1 and P2. In both periods, companies have prioritized environmental aspects such as waste management, emissions, natural resources (water, energy), and biodiversity, reflecting the need to address the inherent challenges of a resource-intensive tourism sector [41]. Additionally, terms related to communication and indicators, such as reports, memory, and information, as well as risk management, stand out, evidencing efforts to improve transparency and strategic planning.

Figure 11. Topics related to CS in P1.
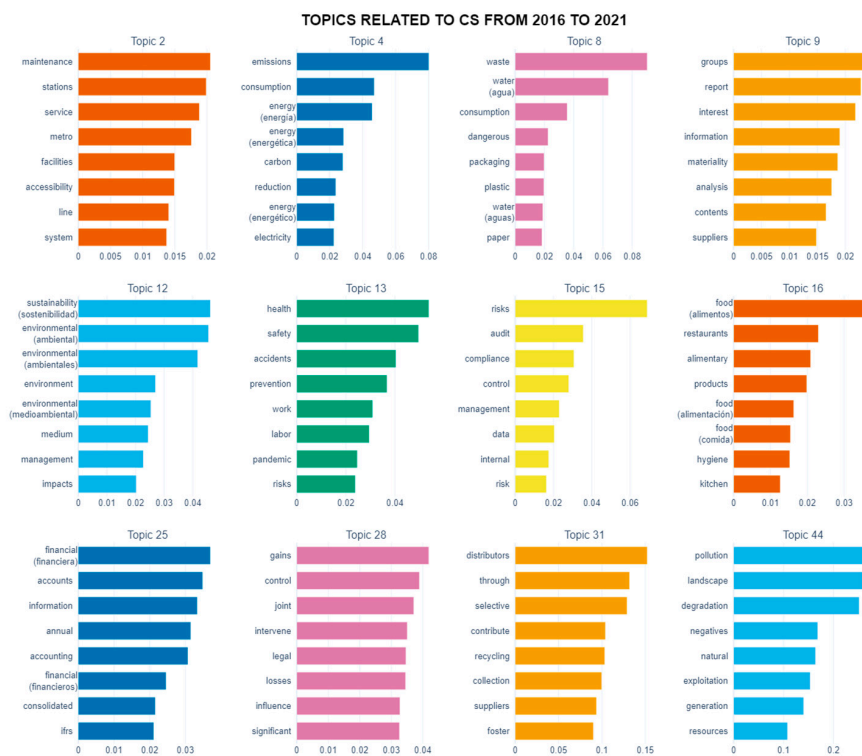
Figure 12. Topics related to CS in P2.

In P1, topics include services, telecommunications, and safety, while in P2, priorities such as carbon footprint reduction, recycling, and accessibility emerge, marking progress toward strategic sustainability that impacts social, environmental, and economic performance [124]. Furthermore, the integration of financial topics, such as annual accounts, in P2 underscores the importance of responsibly managing economic, social, and environmental impacts to foster sustainable development [125]. This analysis reflects a transition in thematic approaches and corporate management priorities between the two periods, adapting to contemporary challenges and consolidating key areas for CS and CSR.

5.2.3. Descriptive Statistical Analysis

Following the established methodology, the relative frequencies of the 364 corporate reports were calculated, distinguishing between CSR and CS categories. Figure 13 presents the box plots corresponding to the annual relative frequencies. During P1, the medians for CSR are higher than those for CS, suggesting a predominant focus on this category. However, starting in P2, the medians for CS surpass those for CSR, reflecting a shift toward themes related to CS.
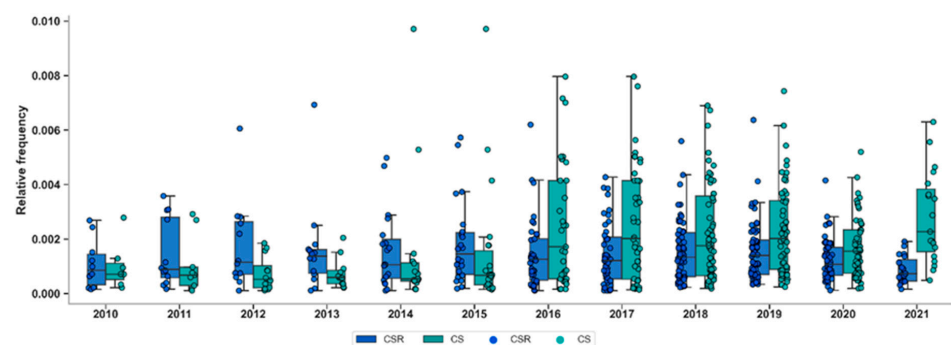

Figure 13. Box plots of the relative frequencies of the reports.

The increase in data point density from 2016 onward indicates a greater number of analyzed reports, especially in the CS category. Additionally, the maximum value reached in 2021 for CS stands out as the highest point of relative frequency in the studied period, which could indicate a significant growth in this category within corporate practices.

In contrast, the relative frequencies for CSR show greater stability over time, with controlled dispersion during the 2016–2021 period. Meanwhile, CS exhibits greater variability in some years, reflecting differences in the representation of this category in the reports.

Figure 14 presents box plots by category and period. In Figure 14a, corresponding to CSR, a higher density of data points is observed in P2 compared to P1, with more concentrated relative frequencies and less dispersion in the second period. In Figure 14b, related to CS, the relative frequencies in P2 exhibit a wider range and a notable increase compared to those in P1, highlighting the evolution of this category.
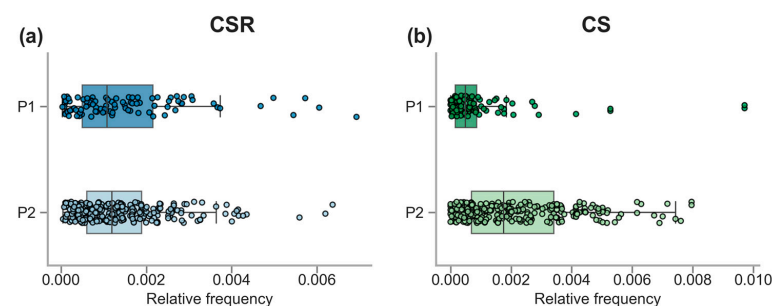

Figure 14. Box plots by category and period. Panel (**a**) represents the corporate social responsibility (CSR) category, while panel (**b**) represents the corporate sustainability (CS) category.

5.2.4. Normality and Bootstrapping Tests

To analyze significant differences between the relative frequencies of CSR and CS categories, normality was initially assessed using the D'Agostino test. This test allowed for testing the null hypothesis that the data do not significantly differ from a normal distribution. The results obtained (p-value < 0.05) confirmed the absence of normality across all categories and periods, justifying the use of non-parametric methods for inferential analysis.

Bootstrapping was chosen for its ability to estimate sampling variability without assuming a specific distribution, overcoming the limitations of traditional parametric methods, such as underestimating variability and generating biased confidence intervals. Its robustness against outliers and asymmetric distributions makes it particularly suitable for the data analyzed. However, this method also has limitations, such as its dependence on the representativeness of the original sample and the assumption of independence between observations. These considerations were carefully evaluated when interpreting the results to ensure that the conclusions were consistent with the nature of the data. Additionally, to reinforce the validity of the analysis, the bootstrapping results were complemented with the calculation of effect sizes (Cohen's d) and 95% confidence intervals.

The distributions generated by bootstrapping for both categories are presented in Figure 15, showing the variations in mean differences based on random sampling.
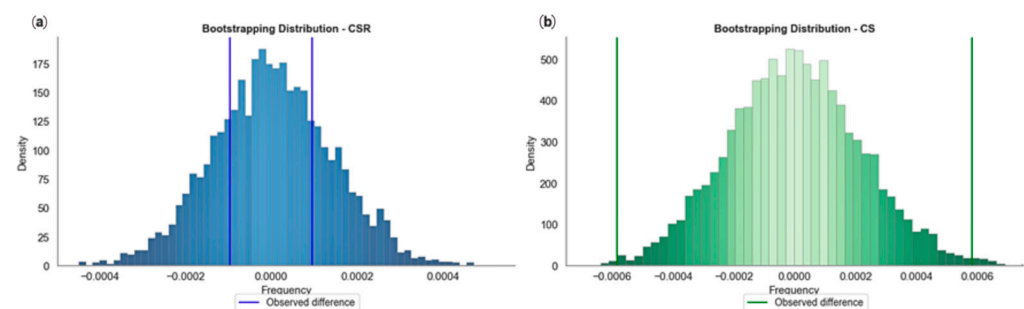

Figure 15. Bootstrapping distributions. Panel (**a**) represents the corporate social responsibility (CSR) category, while panel (**b**) represents the corporate sustainability (CS) category. Each plot includes two vertical lines indicating the observed mean difference.

In the CSR category, the mean in P1 was 0.001491 (with a 95% confidence interval between 0.001198 and 0.001784), while in P2, it was 0.001396 (95% confidence interval between 0.001266 and 0.001525). The observed difference between the means is −0.000095, represented by the blue vertical line in Figure 15a. The bootstrapping analysis produced a p-value > 0.05, indicating no significant differences between the periods. This result is supported by the small and negative effect size (Cohen's d = −0.081), suggesting minimal change in the representation of this category over time. Additionally, the variability (interquartile range, IQR) decreased from 0.001683 in P1 to 0.001294 in P2, which could indicate greater consistency in CSR communication in more recent reports.

On the other hand, in the CS category, the mean in P1 was 0.000940 (with a 95% confidence interval between 0.000612 and 0.001268), significantly increasing to 0.002188 in P2 (95% confidence interval between 0.001971 and 0.002405). The observed difference between the means, shown as the green vertical line in Figure 15b, is 0.001248. The bootstrapping analysis produced a p-value < 0.05, supporting the hypothesis of a significant change in this category. The large effect size (Cohen's d = 0.711) confirms the magnitude of this difference. Additionally, an increase in variability (IQR from 0.000708 to 0.002720) is observed, which could reflect diversification in CS practices during this period.

These results allow for the evaluation of the research hypotheses, as follows:

- H1: "There is a significant change over time in the category 'corporate social responsibility' observed through the words used in corporate reports of tourism companies" is rejected, as no statistically significant differences were observed between P1 and P2.
- H2: "There is a significant change over time in the category 'corporate sustainability' observed through the words used in corporate reports of tourism companies" is accepted, given the significant increase observed in this category.

5.2.5. Importance of ESG Criteria in Spanish Tourism Companies

The results of the text classification using the FinBERT model applied to the corporate reports of tourism companies are presented in Figure 16. The social dimensions stand out as the most representative in the content of the reports, with a significant emphasis on the human capital category (average of 0.21). This focus reflects the prioritization of aspects related to employee well-being, professional development, and working conditions. Complementarily, community relations (average of 0.15) stands out within social strategies, suggesting efforts to strengthen companies' ties with their social environment.
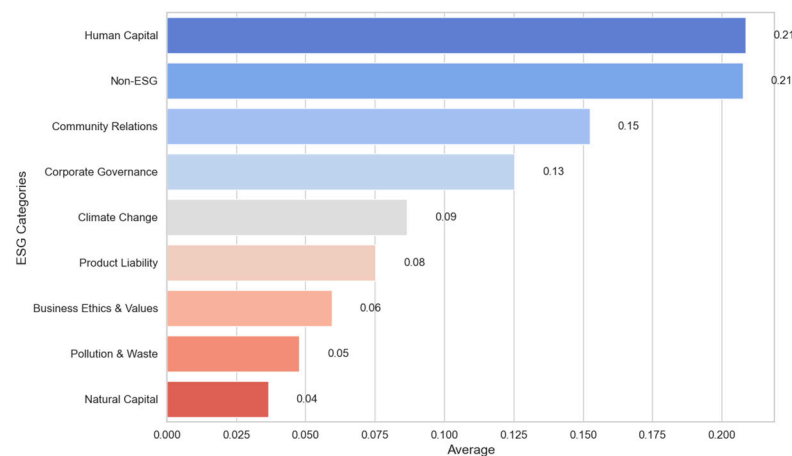

Figure 16. Distribution of ESG (average) categories in corporate reports.

In the corporate governance dimension (average of 0.13), the results highlight the importance of establishing robust organizational structures to ensure transparency, ethics, and responsible decision making. Conversely, environmental dimensions show lower scores, i.e., climate change (0.09), pollution and waste (0.05), and natural capital (0.04). These figures indicate that, while relevant, environmental aspects have not yet reached a priority level comparable to that of social and governance dimensions.

The non-ESG category, with an average equal to that of human capital (0.21), reveals that a significant proportion of the reports address topics outside the ESG criteria. This suggests the coexistence of traditional corporate approaches, alongside efforts toward sustainability, particularly in subsectors facing greater challenges in adopting ESG practices.

Figure 17 illustrates the evolution of ESG categories between 2010 and 2021, highlighting progressive growth across nearly all dimensions. Human capital remains a consistent priority throughout the analyzed period, evidencing the tourism sector's commitment to its workforce. Starting in 2016, a more pronounced increase is observed in categories such as climate change, likely driven by the introduction of stricter European regulations on emissions (e.g., the European Green Deal) and greater social pressure to address global challenges. This growth also coincides with a strengthening in corporate governance and community relations, reflecting a more robust integration of sustainable practices into business strategies.
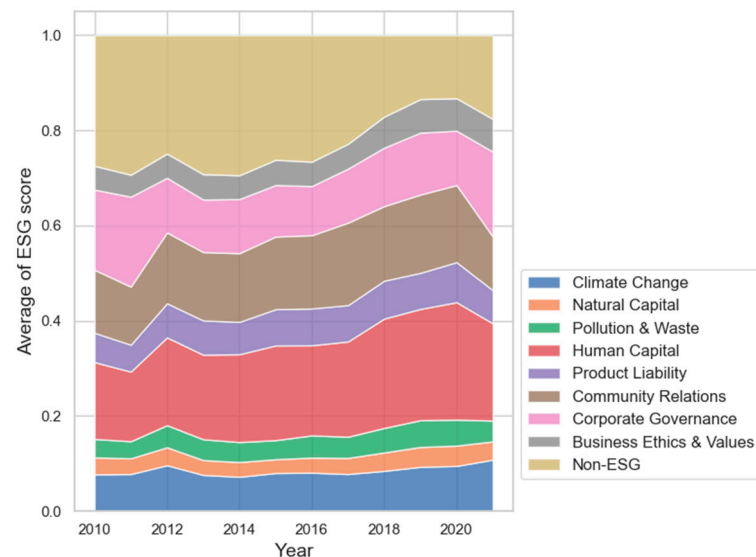
Figure 17. Evolution of ESG categories over time.

When analyzing relationships between ESG categories, possible interdependencies are identified. For instance, the increase in corporate governance may have facilitated the implementation of environmental strategies, such as those related to climate change, and social strategies, such as human capital. These findings suggest that advancements in governance provide a structural foundation for more effectively integrating sustainable practices.

Figure 18 analyzes the average distribution of ESG categories across three subsectors: transport, hotels and accommodations, and restaurants.
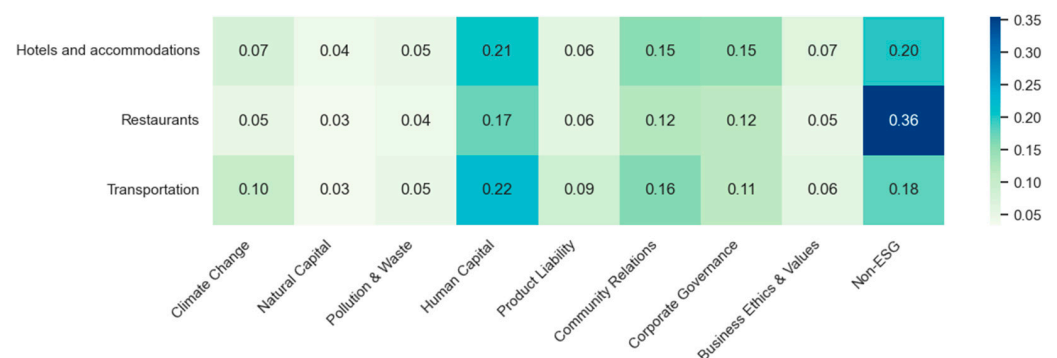

Figure 18. Average of ESG categories by tourism subsectors.

In the transport subsector, environmental categories such as climate change (average of 0.10) stand out due to strict emissions regulations in the air, land, and maritime transport industries. This subsector also shows a strong focus on social categories such as human capital (0.22) and community relations (0.16), highlighting its commitment to employee welfare and the communities affected by its operations.

In hotels and accommodations, the corporate governance category (0.15) stands out, reflecting an emphasis on transparency and ethical management. Similarly, human capital (0.21) and community relations (0.15) emerge as key areas, underscoring the importance of personnel and social interactions in this subsector.

On the other hand, the restaurants subsector presents relatively lower scores in ESG categories, with a notable predominance of the non-ESG category (0.36). This suggests a slower integration of ESG criteria in this subsector, possibly due to fewer financial resources or less stringent regulations compared to those for the other subsectors. However,

gradual improvements have been observed since 2016 in categories such as human capital, community relations, and product responsibility.

Figure 19 illustrates the evolution of ESG scores for tourism subsectors, highlighting differentiated trends based on the specific characteristics of each subsector.
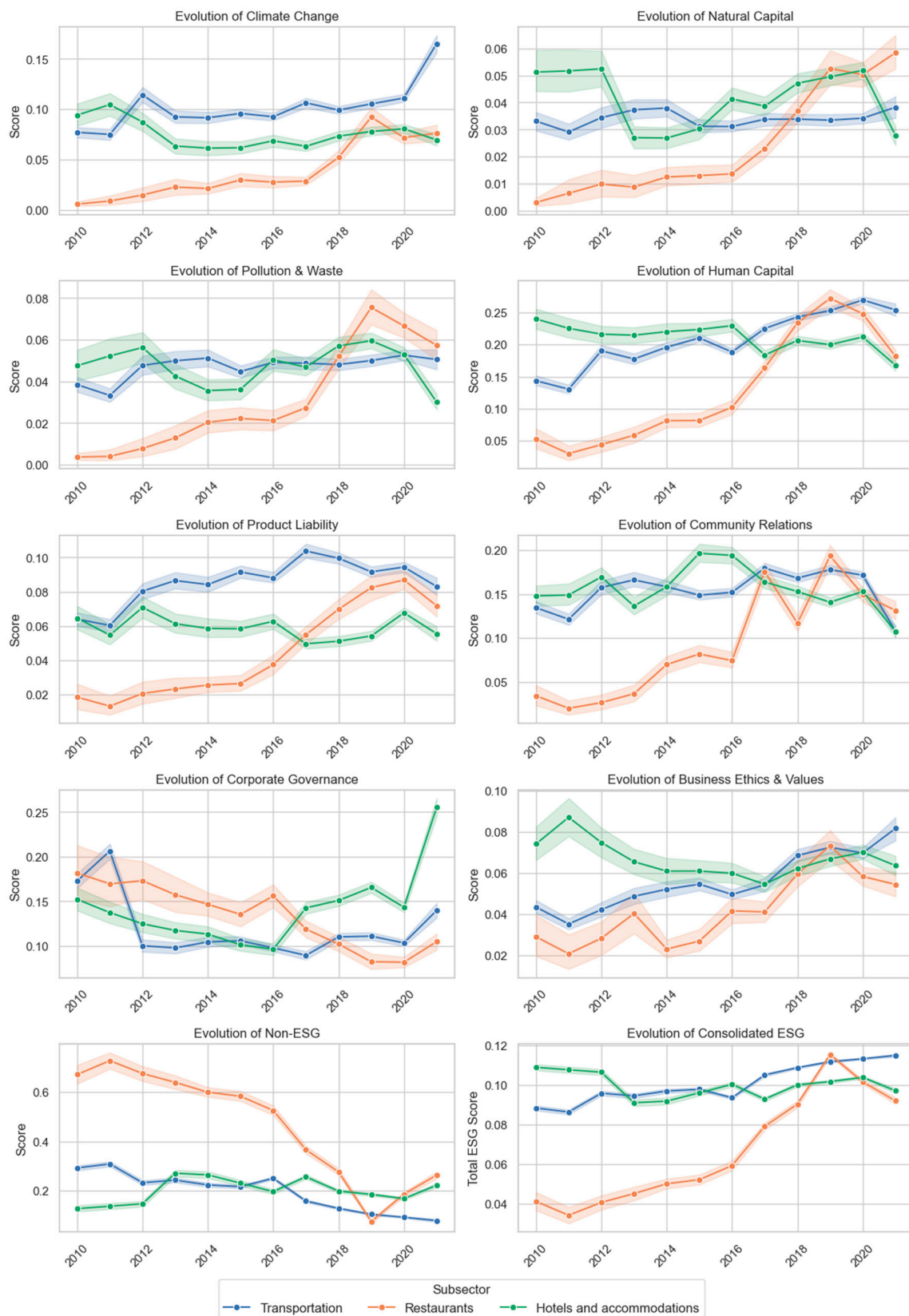

Figure 19. Evolution of each ESG category over time.

In the transport subsector, significant increases are observed in the climate change and human capital categories, reflecting efforts to reduce emissions and improve working conditions. These changes align with policies such as the emissions trading scheme (ETS) and carbon taxes, which have proven to be key tools for mitigating climate change in Europe. According to Zhang [126], the ETS, implemented in 2005 and extended to the aviation sector in 2012, has reduced CO_2 emissions in this sector, albeit with moderate effects on flight costs and demand. Similarly, carbon taxes have incentivized sustainable practices by increasing the costs of land transport and accommodations, promoting the choice of closer and more sustainable destinations.

In the hotels and accommodations subsector, an increase is noted in the corporate governance category, suggesting efforts to improve transparency and ethics in operations. Additionally, social categories such as human capital and community relations maintain high scores, emphasizing sustained attention to employees and local communities. However, Lin et al. [127] point out that this subsector faces significant barriers to adopting ESG standards due to its high labor intensity, heavy reliance on natural and energy resources, and diverse stakeholders. These characteristics require tailored ESG approaches to maximize social, environmental, and economic impacts.

On the other hand, the restaurants subsector shows gradual improvements since 2016 in social categories such as human capital, community relations, and product responsibility. These improvements reflect a growing focus on strategies related to food safety and supply chain management, which are priority issues in the restaurant industry, according to Lee et al. [128]. However, relatively low scores in the environmental and governance categories suggest that this subsector faces specific challenges in adopting sustainable practices.

6. Conclusions

This study examines the theoretical and empirical transition from corporate social responsibility (CSR) to corporate sustainability (CS) in Spanish tourism companies, emphasizing the integration of ESG criteria. To achieve this, various data science techniques and statistical tests are applied, adopting a multidisciplinary approach with a novel and replicable methodology for future studies. This approach helps reduce the gap in the current literature, since despite the significant opportunities offered by data science to tourism research, few studies have explored this direction [16–18]. It also overcomes the limitations of human reading and processing by addressing large volumes of information. Consequently, the findings validate and respond to the proposed research questions and hypotheses, generating both theoretical and practical implications.

6.1. Conclusions Regarding the Theoretical Approach: Computational Literature Review

Theoretically, this research enriches scientific knowledge on CSR and CS in tourism, providing an updated and in-depth perspective through a computational literature review using the topic modeling machine learning technique. A total of 1505 documents from the Scopus database (2004–2023) were analyzed and complemented with statistical analyses. This approach provides clear answers to the first two research questions: 1. What topics have been studied in the existing literature on CSR and CS in tourism over the past twenty years?, and 2. Is there a theoretical paradigm shift from CSR to CS in tourism?

Based on the research topics identified in the literature on CSR and CS, it can be concluded that both approaches are connected to fundamental aspects of tourism. Tourism development emerges as a central theme, reflecting the need to balance economic growth with environmental protection and social aspects. Additionally, a significant shift in research dynamics on CSR and CS associated with the effects of the COVID-19 pandemic

on the sector is evident. Other noteworthy topics include the hotel industry, customer loyalty, and human resource management, demonstrating a more comprehensive focus on different stakeholder groups.

Furthermore, the statistical results establish a theoretical evolution from CSR to CS. This shift occurs gradually. Research on CS has shown steady and significant growth, especially in recent years, while that regarding CSR has stabilized or even slightly declined, though not significantly. This indicates that scientific dynamics in the tourism field are evolving toward CS, supporting the hierarchical relationship between CSR and CS proposed by Van Marrewijk [20], in which CSR serves as an intermediate step toward the ultimate goal of CS.

The results enrich the literature, as a solid theoretical understanding can promote coherent changes that reduce ambiguity in academia and business practices [15].

6.2. Conclusions Regarding the Empirical Approach: Review of Corporate Reports of Tourism Companies in Spain

The empirical analysis provides a comprehensive view of how companies in the Spanish tourism sector have evolved in implementing management practices associated with CSR and CS through a detailed examination of 364 corporate reports published between 2010 and 2021. This research is particularly notable, as it focuses on companies that use Spanish as their corporate language, representing a significant contribution to scientific knowledge in a field dominated by studies conducted in English-speaking contexts [129]. The practical implications derived from this study are diverse.

First, the study enhances the understanding of the key management areas associated with CSR and CS in Spanish tourism companies, addressing the third research question: What are the key areas related to the management of CSR and CS in the corporate reports of tourism companies? The results show that the tourism industry has progressively implemented practices related to organizational management, ethics, equality, diversity, inclusion, and key environmental aspects (such as waste, emissions, and natural resources). In more recent years (2016–2021), additional areas such as carbon footprint, risk management, and recycling have emerged. A stronger connection between financial topics, accounting, and annual accounts with CS is also observed, opening opportunities for future research on the impact of sustainability on financial performance.

Previous studies, such as those by Rhou and Singal [130] and Kaur et al. [131], support this line of inquiry by highlighting inconsistencies in financial results related to CSR but suggesting potential benefits in areas such as cost reduction, customer satisfaction, and organizational commitment. These findings underscore the need to explore metrics such as return on assets (ROA) and the cost of capital under sustainable approaches.

Secondly, the paradigm shift toward CS is statistically validated through the analysis of term usage in corporate reports, addressing the proposed research hypotheses: H1: There is a significant change over time in the "corporate social responsibility" category observed through the words used in corporate reports of tourism companies; and H2: There is a significant change over time in the "corporate sustainability" category observed through the words used in corporate reports of tourism companies. The results confirm significant growth in the presence of terms associated with CS, while those associated with CSR show a stable or slightly declining trend.

This shift, although statistically significant only for CS, indicates that companies continue to use CSR as a central management strategy but are evolving toward a more holistic, sustainability-focused approach. This finding aligns with the literature positioning CS as a more comprehensive framework in business strategies [7,8]. In this process, ESG criteria have emerged as a solid standard for evaluating and measuring CS, as they incorporate stakeholder expectations and demands [31,32].

The transition toward CS is driven by the ESG approach, which not only promotes the implementation of sustainable practices but also creates business value. Studies have shown that companies adopting ESG criteria strengthen their profitability, crisis resilience, and risk management capacity [36–38]. Moreover, integrating these principles enhances corporate perception, increasing customer satisfaction and brand loyalty [7].

Thirdly, ESG criteria are analyzed, addressing the fourth research question: What importance do tourism companies and subsectors attribute to ESG criteria in the corporate reports of the Spanish tourism sector? The results reveal differences between subsectors in how they address ESG criteria. The transport sector stands out as the most proactive in terms of environmental sustainability, with a strong focus on climate change and product responsibility, driven by strict regulations and social demands [132]. However, it is notable that this subsector needs to improve governance, as Sim and Kim [133] emphasize that the governance dimension is key to strengthening brand perception, emotional attachment, and consumer behavioral intentions.

In the field of hotels and accommodations, corporate governance and community relations are priorities, reflecting a strategy aimed at improving transparency and strengthening social responsibility. These attributes are crucial in this subsector, as they influence customer perception, premium pricing, and brand choice. Yu et al. [36] identify key ESG attributes in environmental, social, and governance dimensions, highlighting, among others, natural resource conservation, corporate ethics, diversity, and community engagement.

On the other hand, the restaurant subsector demonstrates a more limited integration of ESG criteria, possibly due to barriers such as high initial costs and fragmented supply chains, as noted by Madanaguli et al. [134]. However, sustainable practices such as waste reduction, the use of organic products, and efficient water management stand out, driven by consumer pressure and government regulations. To overcome these limitations, it would be beneficial to develop fiscal incentives and sustainability training programs that enable companies to adopt ESG practices more efficiently.

These findings highlight the areas where companies have made progress and identify those where further integration of sustainable practices is needed, providing useful and critical information to guide both business decision making and public policy formulation in the future. Similarly, they enrich longitudinal and sectoral studies regarding CSR and CS, reducing the current gap in the literature [7,19].

6.3. Limitations and Future Lines of Research

This study presents methodological and scope-related limitations that should be considered when interpreting its results. Theoretically, the literature review relies exclusively on the Scopus database, which may exclude relevant contributions indexed in other academic sources. Empirically, the analysis focuses solely on the Spanish tourism sector and corporate reports in Spanish, potentially limiting the extrapolation of results to other sectors or international contexts. To address these limitations, future studies could expand the theoretical scope by utilizing additional databases, such as Web of Science. Similarly, incorporating sustainability reports submitted to international organizations, such as the Global Reporting Initiative (GRI), would facilitate standardized comparisons between companies from different countries and cultural contexts, providing a more robust framework for the global analysis of CS.

Furthermore, the application of artificial intelligence and advanced machine learning techniques in the analysis of sustainability reports represents a key opportunity for future research. More sophisticated natural language processing (NLP) models could enhance the identification of patterns in the evolution of corporate discourse on CS and ESG, enabling a more automated and precise evaluation, particularly when analyzing multilingual

documents. The integration of these approaches would contribute to a more detailed and comparative analysis of corporate reports, optimizing the extraction of relevant information and facilitating broader studies.

On the other hand, the sample includes only large and medium-sized organizations, excluding the perspective of small and microenterprises, which constitute a significant portion of the tourism business ecosystem. Future studies could develop methodological frameworks specifically designed to evaluate CSR and CS practices in small businesses, integrating both financial and non-financial metrics. Complementing text analysis with qualitative approaches, such as in-depth interviews and content analysis of corporate websites and social media, could provide a more comprehensive understanding of how these organizations communicate and manage sustainability in dynamic environments.

Other future research directions could focus on analyzing sustainability certifications, such as Tripadvisor's green hotel programs, and their inclusion in the corporate reports of tourism companies. This would allow for the evaluation of how these certifications reflect a commitment to sustainable practices and impact reputation and management strategies. Additionally, exploring companies' alignment with the Sustainable Development Goals (SDGs), particularly SDG 8 (decent work and sustainable tourism), SDG 12 (responsible production and consumption), and SDG 13 (climate action), would be valuable. The use of advanced data science techniques, such as topic modeling and natural language processing, could facilitate the analysis of textual data, identifying patterns and opportunities in the adoption of these initiatives.

Author Contributions: Conceptualization, M.F.B.S., E.B.-D. and J.C.-A.; Methodology, M.F.B.S., E.B.-D. and J.C.-A.; Software, M.F.B.S.; Investigation, M.F.B.S.; Data curation, M.F.B.S.; Writing—original draft, M.F.B.S.; Writing—review & editing, M.F.B.S., E.B.-D. and J.C.-A.; Visualization, M.F.B.S.; Supervision, E.B.-D. and J.C.-A.; Project administration, E.B.-D. and J.C.-A. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

1. Lee, S.; Yeon, J.; Song, H.J. Current status and future perspective of the link of corporate social responsibility–corporate financial performance in the tourism and hospitality industry. *Tour. Econ.* **2022**, *29*, 1703–1735. [CrossRef]
2. Font, X.; Walmsley, A.; Cogotti, S.; McCombes, L.; Häusler, N. Corporate social responsibility: The disclosure-performance gap. *Tour. Manag.* **2012**, *33*, 1544–1553. [CrossRef]
3. Font, X.; Lynes, J. Corporate social responsibility in tourism and hospitality. *J. Sustain. Tour.* **2018**, *26*, 1027–1042. [CrossRef]
4. Montiel, I. Corporate social responsibility and corporate sustainability: Separate pasts, common futures. *Organ. Environ.* **2008**, *21*, 245–269. [CrossRef]
5. Sheehy, B.; Farneti, F. Corporate social responsibility, sustainability, sustainable development and corporate sustainability: What is the difference, and does it matter? *Sustainability* **2021**, *13*, 5965. [CrossRef]
6. Lu, H. The 'legalisation' of corporate social responsibility: Hong Kong experience on ESG reporting. *Asia Pac. Law Rev.* **2016**, *24*, 123–148. [CrossRef]

7. Back, K.J. ESG for the hospitality and tourism research: Essential demanded research area for all. *Tour. Manag.* **2024**, *105*, 104954. [CrossRef]

8. Legendre, T.S.; Ding, A.; Back, K.J. A bibliometric analysis of the hospitality and tourism environmental, social, and governance (ESG) literature. *J. Hosp. Tour. Manag.* **2024**, *58*, 309–321. [CrossRef]

9. Grosbois, D. Corporate social responsibility reporting by the global hotel industry: Commitment, initiatives and performance. *Int. J. Hosp. Manag.* **2012**, *31*, 896–905. [CrossRef]

10. Medrado, L.; Jackson, L.A. Corporate nonfinancial disclosures: An illuminating look at the corporate social responsibility and sustainability reporting practices of hospitality and tourism firms. *Tour. Hosp. Res.* **2015**, *16*, 116–132. [CrossRef]

11. Diwan, H.; Amarayil Sreeraman, B. From financial reporting to ESG reporting: A bibliometric analysis of the evolution in corporate sustainability disclosures. *Environ. Dev. Sustain.* **2023**, *26*, 13769–13805. [CrossRef]

12. Uyar, A.; Köseoglu, M.A.; Kılıç Karamahmutoğlu, M.; Mehraliyev, F. Thematic structure of sustainability reports of the hospitality and tourism sector: A periodical, regional, and format-based analysis. *Curr. Issues Tour.* **2020**, *24*, 2602–2627. [CrossRef]

13. Mastercard; Turespaña. Expectativas y Comportamiento de los Viajeros Europeos en Materia de Sostenibilidad. 2023. Available online: https://www.tourspain.es/es/notas-prensa-turismo/espana-mejores-destinos-turisticos-viajeros-sostenibles/ (accessed on 25 September 2024).

14. Google; Deloitte. Viajeros y Destinos NextGen: Nuestra Visión Sobre la Transformación del Sector Turístico. 2024. Available online: https://perspectivas.deloitte.com/l/915781/2024-06-11/t9wxf/915781/17180970624JfWZbST/20240305_Viajeros_Destinos_NextGen_Google.pdf (accessed on 25 September 2024).

15. Miller, G.; Twining-Ward, L. *Monitoring for a Sustainable Tourism Transition: The Challenge of Developing and Using Indicators*; CABI Publishing: Wallingford, UK, 2005. [CrossRef]

16. Egger, R. Applied Data Science in Tourism: Interdisciplinary Approaches, Methodologies, and Applications. In *Tourism on the Verge*; Springer: Cham, Switzerland, 2022.

17. Mariani, M.; Baggio, R. Big data and analytics in hospitality and tourism: A systematic literature review. *Int. J. Contemp. Hosp. Manag.* **2022**, *34*, 231–278. [CrossRef]

18. Mariani, M.; Baggio, R.; Fuchs, M.; Höepken, W. Business intelligence and big data in hospitality and tourism: A systematic literature review. *Int. J. Contemp. Hosp. Manag.* **2018**, *30*, 3514–3554. [CrossRef]

19. Koh, Y. The industry-academia gap on the corporate governance issues in tourism and hospitality: A critical review and future research suggestions. *J. Travel Tour. Mark.* **2024**, *41*, 3–19. [CrossRef]

20. Van Marrewijk, M. Concepts and definitions of CSR and corporate sustainability: Between agency and communion. *J. Bus. Ethics* **2003**, *44*, 95–105. [CrossRef]

21. Lechuga Sancho, M.P.; Martín-Navarro, A.; Ramos-Rodríguez, A.R. Information Systems Management Tools: An Application of Bibliometrics to CSR in the Tourism Sector. *Sustainability* **2020**, *12*, 8697. [CrossRef]

22. Madanaguli, A.; Srivastava, S.; Ferraris, A.; Dhir, A. Corporate social responsibility and sustainability in the tourism sector: A systematic literature review and future outlook. *Sustain. Dev.* **2022**, *30*, 447–461. [CrossRef]

23. Aragon-Correa, J.A.; Martin-Tapia, I.; de la Torre-Ruiz, J. Sustainability issues and hospitality and tourism firms' strategies: Analytical review and future directions. *Int. J. Contemp. Hosp. Manag.* **2015**, *27*, 498–522. [CrossRef]

24. Sánchez-Camacho, C.; Carranza, R.; Martín-Consuegra, D.; Díaz, E. Evolution, trends and future research lines in corporate social responsibility and tourism: A bibliometric analysis and science mapping. *Sustain. Dev.* **2022**, *30*, 462–476. [CrossRef]

25. Wut, T.M.; Xu, J.; Wong, S.M.H. Timing and congruence effects of corporate social responsibility practices on social media crises in the tourism industry. *J. Hosp. Tour. Technol.* **2023**, *14*, 154–171. [CrossRef]

26. Martínez, P.; Rodríguez del Bosque, I. Sustainability Dimensions: A Source to Enhance Corporate Reputation. *Corp. Reput. Rev.* **2014**, *17*, 239–253. [CrossRef]

27. Martínez, P.; Rodríguez del Bosque, I. CSR and customer loyalty: The roles of trust, customer identification with the company and satisfaction. *Int. J. Hosp. Manag.* **2013**, *35*, 89–99. [CrossRef]

28. Yoopetch, C.; Nimsai, S.; Kongarchapatara, B. Bibliometric Analysis of Corporate Social Responsibility in Tourism. *Sustainability* **2023**, *15*, 668. [CrossRef]

29. Tuan, L.T.; Rajendran, D.; Rowley, C.; Khai, D.C. Customer value co-creation in the business-to-business tourism context: The roles of corporate social responsibility and customer empowering behaviors. *J. Hosp. Tour. Manag.* **2019**, *39*, 137–149. [CrossRef]

30. Andreu Pinillos, A.; Fernández Fernández, J.L. De la RSC a la sostenibilidad corporativa: Una evolución necesaria para la creación de valor. *Harv. Deusto Bus. Rev.* **2011**, *207*, 5–21.

31. Burbano, V.C.; Delmas, M.A.; Cobo, M.J. The Past and Future of Corporate Sustainability Research. *Organ. Environ.* **2023**, *37*, 133–158. [CrossRef]

32. Serafeim, G. Social-impact efforts that create real value. *Harv. Bus. Rev.* **2020**, 1–20.

33. David, L.K.; Wang, J.; Angel, V.; Luo, M. China's ESG scorecard: A predictive machine learning model. *Corp. Soc. Responsib. Environ. Manag.* **2024**, *31*, 3468–3486. [CrossRef]

34. Hanson, D.; Usa, J.F.; Koller, T.; Jiang, B.; Raj, R.; Tarasoff, J.; Capital, G.L.; Mccormack, J.; Park, A.; Ravenel, C.; et al. ESG Investing in Graham & Doddsville. *J. Appl. Corp. Financ.* **2013**, *25*, 20–31. [CrossRef]

35. Kumar, V.; Srivastava, A. Trends in the thematic landscape of corporate social responsibility research: A structural topic modeling approach. *J. Bus. Res.* **2022**, *150*, 26–37. [CrossRef]

36. Yu, J.; Chiriko, A.Y.; Kim, S.; Moon, H.G.; Choi, H.; Han, H. ESG management of hotel brands: A management strategy for benefits and performance. *Int. J. Hosp. Manag.* **2025**, *125*, 103998. [CrossRef]

37. Zhang, D.; Liu, L. Does ESG Performance Enhance Financial Flexibility? Evidence from China. *Sustainability* **2022**, *14*, 11324. [CrossRef]

38. Chen, C.D.; Su, C.H.J.; Chen, M.H. Are ESG-committed hotels financially resilient to the COVID-19 pandemic? An autoregressive jump intensity trend model. *Tour. Manag.* **2022**, *93*, 104581. [CrossRef] [PubMed]

39. Cohen, R.; Kennedy, P. *Global Sociology*; Bloomsbury Publishing: London, UK, 2017.

40. Henderson, J.C. Corporate social responsibility and tourism: Hotel companies in Phuket, Thailand, after the Indian Ocean tsunami. *Int. J. Hosp. Manag.* **2007**, *26*, 228–239. [CrossRef]

41. Gössling, S.; Peeters, P. Assessing tourism's global environmental impact 1900–2050. *J. Sustain. Tour.* **2025**, *23*, 639–659. [CrossRef]

42. Sørensen, F.; Bærenholdt, J.O. Tourist practices in the circular economy. *Ann. Tour. Res.* **2020**, *85*, 103027. [CrossRef]

43. Strippoli, R.; Gallucci, T.; Ingrao, C. Circular economy and sustainable development in the tourism sector—An overview of the truly-effective strategies and related benefits. *Heliyon* **2024**, *10*, e36801. [CrossRef]

44. Gül, H.; Çağatay, S.; Gartner, W.C. Fossil fuel and resource use impacts of raw material substitution and recycling in tourism sector in Türkiye: Evidence from circular economy adjusted input-output matrix. *Curr. Issues Tour.* **2024**, 1–20. [CrossRef]

45. García, C.; Herrero, B.; Morillas-Jurado, F. The impact of the environmental, social and governance dimensions of sustainability on firm risk in the hospitality and tourism industry. *Corp. Soc. Responsib. Environ. Manag.* **2024**, *31*, 2783–2800. [CrossRef]

46. Alonso-Muñoz, S.; Torrejón-Ramos, M.; Medina-Salgado, M.S.; González-Sánchez, R. Sustainability as a building block for tourism—Future research: Tourism Agenda 2030. *Tour. Rev.* **2023**, *78*, 461–474. [CrossRef]

47. Sharma, A.; Nunkoo, R.; Rana, N.P.; Dwivedi, Y.K. On the intellectual structure and influence of tourism social science research. *Ann. Tour. Res.* **2021**, *91*, 103142. [CrossRef]

48. Nunkoo, R.; Sharma, A.; Rana, N.P.; Dwivedi, Y.K.; Sunnassee, V.A. Advancing sustainable development goals through interdisciplinarity in sustainable tourism research. *J. Sustain. Tour.* **2023**, *31*, 735–759. [CrossRef]

49. Rasoolimanesh, S.M.; Ramakrishna, S.; Hall, C.M.; Esfandiar, K.; Seyfi, S. A systematic scoping review of sustainable tourism indicators in relation to the sustainable development goals. *J. Sustain. Tour.* **2023**, *31*, 1497–1517. [CrossRef]

50. Girella, L.; Zambon, S.; Rossi, P. Reporting on sustainable development: A comparison of three Italian small and medium-sized enterprises. *Corp. Soc. Responsib. Environ. Manag.* **2019**, *26*, 981–996. [CrossRef]

51. Miralles-Quiros, M.D.M.; Miralles-Quiros, J.L.; Arraiano, I.G. Are Firms that Contribute to Sustainable Development Valued by Investors? *Corp. Soc. Responsib. Environ. Manag.* **2017**, *24*, 71–84. [CrossRef]

52. Hsu, C.W.; Lee, W.H.; Chao, W.C. Materiality analysis model in sustainability reporting: A case study at Lite-On Technology Corporation. *J. Clean. Prod.* **2013**, *57*, 142–151. [CrossRef]

53. Kumar, D. Role of corporate sustainability disclosures in moderating the impact of country-level uncertainties on tourism sector firms' risk. *J. Sustain. Tour.* **2024**, *32*, 1287–1306. [CrossRef]

54. Johann, M. CSR Strategy in Tourism during the COVID-19 Pandemic. *Sustainability* **2022**, *14*, 3773. [CrossRef]

55. Zientara, P.; Adamska, J.; Bąk, M. What do tourism and hospitality companies convey about labor union relations in their CSR reports? A conceptual model and empirical findings. *J. Sustain. Tour.* **2024**, 1–24. [CrossRef]

56. Gajjar, T.; Okumus, F. Diversity management: What are the leading hospitality and tourism companies reporting? *J. Hosp. Mark. Manag.* **2018**, *27*, 905–925. [CrossRef]

57. Quintás, M.A.; Martínez-Senra, A.I.; García-Pintos, A. Is the hotel industry really committed to the environment? Answering using the business models framework. *Serv. Bus.* **2023**, *17*, 395–428. [CrossRef]

58. Geerts, M.; Dooms, M. An analysis of the CSR portfolio of cruise shipping lines. *Res. Transp. Bus. Manag.* **2022**, *45*, 100615. [CrossRef]

59. Di Vaio, A.; Varriale, L.; Lekakou, M.; Pozzoli, M. SDGs disclosure: Evidence from cruise corporations' sustainability reporting. *Corp. Gov.* **2023**, *23*, 845–866. [CrossRef]

60. Di Vaio, A.; Hassan, R.; D'amore, G.; Dello Strologo, A. Digital Technologies for Sustainable Waste Management On-Board Ships: An Analysis of Best Practices from the Cruise Industry. *IEEE Trans. Eng. Manag.* **2024**, *71*, 12715–12728. [CrossRef]

61. Yoon, J.; Han, S.; Lee, Y.; Hwang, H. Text Mining Analysis of ESG Management Reports in South Korea: Comparison with Sustainable Development Goals. *SAGE Open* **2023**, *13*, 21582440231202896. [CrossRef]

62. Antons, D.; Breidbach, C.F.; Joshi, A.M.; Salge, T.O. Computational Literature Reviews: Method, Algorithms, and Roadmap. *Organ. Res. Methods* **2021**, *26*, 107–138. [CrossRef]

63. Mortenson, M.J.; Vidgen, R. A computational literature review of the technology acceptance model. *Int. J. Inf. Manag.* **2016**, *36*, 1248–1259. [CrossRef]

64. Grootendorst, M. MaartenGr/BERTopic: Fix Embedding Parameter. 2021. Available online: https://zenodo.org/records/4430182 (accessed on 6 August 2023).

65. Deerwester, S.; Dumais, S.T.; Furnas, G.W.; Landauer, T.K.; Harshman, R. Indexing by latent semantic analysis. *J. Am. Soc. Inf. Sci.* **1990**, *41*, 391–407.

66. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.

67. Wallach, H.M. Topic Modeling: Beyond Bag-of-Words. In Proceedings of the ICML '06: 23rd International Conference on Machine Learning, Pittsburgh, PA, USA, 25–29 June 2006; Association for Computing Machinery: New York, NY, USA, 2006; pp. 977–984.

68. Grootendorst, M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv* **2022**, arXiv:2203.05794. [CrossRef]

69. Uysal, A.K.; Gunal, S. The impact of preprocessing on text classification. *Inf. Process. Manag.* **2014**, *50*, 104–112. [CrossRef]

70. Wang, Z.; Chen, J.; Chen, J.; Chen, H. Identifying interdisciplinary topics and their evolution based on BERTopic. *Scientometrics* **2023**, *129*, 7359–7384. [CrossRef]

71. Douglas, T.; Capra, L.; Musolesi, M. A Computational Linguistic Approach to Study Border Theory at Scale. In Proceedings of the ACM on Human-Computer Interaction, 8(CSCW1), New York, NY, USA, 23 April 2024. [CrossRef]

72. Kim, K.; Kogler, D.F.; Maliphol, S. Identifying interdisciplinary emergence in the science of science: Combination of network analysis and BERTopic. *Humanit. Soc. Sci. Commun.* **2024**, *11*, 1–15. [CrossRef]

73. Kendall, M.G. *Rank Correlation Methods*, 4th ed.; Charles Griffin: London, UK, 1975.

74. Mann, H.B. Nonparametric Tests Against Trend. *Econometrica* **1945**, *13*, 245. [CrossRef]

75. Ning, X.; Yim, D.; Khuntia, J. Online sustainability reporting and firm performance: Lessons learned from text mining. *Sustainability* **2021**, *13*, 1069. [CrossRef]

76. Landrum, N.E.; Ohsowski, B. Identifying Worldviews on Corporate Sustainability: A Content Analysis of Corporate Sustainability Reports. *Bus. Strategy Environ.* **2018**, *27*, 128–151. [CrossRef]

77. Einwiller, S.A.; Carroll, C.E. Negative disclosures in corporate social responsibility reporting. *Corp. Commun.* **2020**, *25*, 319–337. [CrossRef]

78. Ellili, N.O.D.; Nobanee, H. Impact of economic, environmental, and corporate social responsibility reporting on financial performance of UAE banks. *Environ. Dev. Sustain.* **2022**, *25*, 3967–3983. [CrossRef]

79. Torelli, R.; Balluchi, F.; Furlotti, K. The materiality assessment and stakeholder engagement: A content analysis of sustainability reports. *Corp. Soc. Responsib. Environ. Manag.* **2020**, *27*, 470–484. [CrossRef]

80. Asif, M.; Searcy, C.; dos Santos, P.; Kensah, D. A Review of Dutch Corporate Sustainable Development Reports. *Corp. Soc. Responsib. Environ. Manag.* **2013**, *20*, 321–339. [CrossRef]

81. Campopiano, G.; de Massis, A. Corporate Social Responsibility Reporting: A Content Analysis in Family and Non-family Firms. *J. Bus. Ethics* **2015**, *129*, 511–534. [CrossRef]

82. Gatti, L.; Seele, P. Evidence for the prevalence of the sustainability concept in European corporate responsibility reporting. *Sustain. Sci.* **2014**, *9*, 89–102. [CrossRef]

83. SABI. Sistema de Análisis de Balances Ibéricos. Último Acceso a La Web. 2022. Available online: https://sabi.informa.es/ (accessed on 6 August 2023).

84. Branco, M.C.; Rodrigues, L.L. Factors Influencing Social Responsibility Disclosure by Portuguese Companies. *J. Bus. Ethics* **2008**, *83*, 685–701. [CrossRef]

85. Reverte, C. Determinants of corporate social responsibility disclosure ratings by Spanish listed firms. *J. Bus. Ethics* **2009**, *88*, 351–366. [CrossRef]

86. Khan, H. The effect of corporate governance elements on corporate social responsibility (CSR) reporting: Empirical evidence from private commercial banks of Bangladesh. *Int. J. Law Manag.* **2010**, *52*, 82–109.

87. Ministerio de Trabajo y Economía Social. El Portal de la Responsabilidad Social. 2022. Available online: https://www.mites.gob.es/es/rse/ (accessed on 6 August 2023).

88. Pizzi, S.; Caputo, A.; Corvino, A.; Venturelli, A. Management research and the UN sustainable development goals (SDGs): A bibliometric investigation and systematic review. *J. Clean. Prod.* **2020**, *276*, 124033. [CrossRef]

89. Rosato, P.F.; Caputo, A.; Valente, D.; Pizzi, S. 2030 Agenda and Sustainable Business Models in Tourism: A Bibliometric Analysis. In *Ecological Indicators*; Elsevier: Amsterdam, The Netherlands, 2021; Volume 121. [CrossRef]

90. Witten, I.; Frank, E.; Hall, M.A. *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed.; Morgan Kaufmann: Burlington, MA, USA, 2011.

91. Huang, A.; Wang, H.; Yang, Y. FinBERT: A Large Language Model for Extracting Information from Financial Text. *Contemp. Account. Res.* **2022**, *40*, 806–841. [CrossRef]

92. Vives Varela, T.; Hamui Sutton, L. La codificación y categorización en la teoría fundamentada, un método para el análisis de los datos cualitativos. *Investig. Educ. Médica* **2021**, *10*, 97–104. [CrossRef]

93. Karen, P. Corporate sustainability, citizenship and social responsibility reporting: A website study of 100 model corporations. *J. Corp. Citizsh.* **2008**, *32*, 63–78.

94. Palazzo, G.; Scherer, A.G. Corporate legitimacy as deliberation: A communicative framework. *J. Bus. Ethics* **2006**, *66*, 71–88. [CrossRef]

95. UNESCO. Tesauro de la UNESCO. 2022. Available online: https://vocabularies.unesco.org/ (accessed on 6 August 2023).

96. García, S.; Ramírez-Gallego, S.; Luengo, J.; Herrera, F. Big Data: Preprocesamiento y calidad de datos. *Novática* **2016**, *237*, 17–23.

97. Muller, C. *Estadística Lingüística*; Gredos S.A.: Madrid, Spain, 1973.

98. Salinas, C.M.; Martínez, G.E.S. Hacia una normalización de la frecuencia de los corpus CREA y CORDE. *Rev. Signos* **2015**, *48*, 307–331. [CrossRef]

99. D'Agostino, R.; Pearson, E.S. Tests for Departure from Normality. Empirical Results for the Distributions of b2 and $\sqrt{b1}$. *Biometrika* **1973**, *60*, 613–622. [CrossRef]

100. Efron, B. Bootstrap methods: Another look at the Jackknife. *Ann. Stat.* **1979**, *7*, 1–26. [CrossRef]

101. Srijiranon, K.; Lertratanakham, Y.; Tanantong, T. A Hybrid Framework Using PCA, EMD and LSTM Methods for Stock Market Price Prediction with Sentiment Analysis. *Appl. Sci.* **2022**, *12*, 10823. [CrossRef]

102. Sinha, A.; Kedas, S.; Kumar, R.; Malo, P. SEntFiN 1.0: Entity-aware sentiment analysis for financial news. *J. Assoc. Inf. Sci. Technol.* **2022**, *73*, 1314–1335. [CrossRef]

103. Gao, Y.; Zheng, W.; Wang, Y. Sectoral risk contagion and quantile network connectedness on Chinese stock sectors after the COVID-19 outbreak. *China Financ. Rev. Int.* **2023**, *in press*. [CrossRef]

104. Arami, M.; Balina, S.; Chang, V.; Kim, J.; Kim, H.-S.; Choi, S.-Y. Forecasting the S&P 500 Index Using Mathematical-Based Sentiment Analysis and Deep Learning Models: A FinBERT Transformer Model and LSTM. *Axioms* **2023**, *12*, 835. [CrossRef]

105. Xiao, Q.; Ihnaini, B. Stock trend prediction using sentiment analysis. *PeerJ Comput. Sci.* **2023**, *9*, e1293. [CrossRef]

106. Unlu, A.; Truong, S.; Tammi, T.; Lohiniva, A.L. Exploring Political Mistrust in Pandemic Risk Communication: Mixed-Method Study Using Social Media Data Analysis. *J. Med. Internet Res.* **2023**, *25*, e50199. [CrossRef]

107. Unlu, A.; Truong, S.; Sawhney, N.; Sivelä, J.; Tammi, T. Tracing the dynamics of misinformation and vaccine stance in Finland amid COVID-19. *Inf. Commun. Soc.* **2024**, *28*, 193–217. [CrossRef]

108. Fedorova, E.; Iasakova, P. The impact of climate change news on the US stock market. *J. Risk Financ.* **2024**, *25*, 293–320. [CrossRef]

109. Sahu, A.K.; Debata, B.; Dash, S.R. Manager sentiment, policy uncertainty, ESG disclosure and firm performance: A large language model in corporate landscape. *Int. J. Account. Inf. Manag.* **2024**, *32*, 858–882. [CrossRef]

110. Lane, B.; Kastenholz, E. Rural tourism: The evolution of practice and research approaches—Towards a new generation concept? *J. Sustain. Tour.* **2015**, *23*, 1133–1156. [CrossRef]

111. Sigala, M. Tourism and COVID-19: Impacts and implications for advancing and resetting industry and research. *J. Bus. Res.* **2020**, *117*, 312–321. [CrossRef] [PubMed]

112. Budeanu, A.; Miller, G.; Moscardo, G.; Ooi, C.S. Sustainable tourism, progress, challenges and opportunities: An introduction. *J. Clean. Prod.* **2016**, *111*, 285–294. [CrossRef]

113. Cannas, R.; Argiolas, G.; Cabiddu, F. Fostering corporate sustainability in tourism management through social values within collective value co-creation processes. *J. Sustain. Tour.* **2019**, *27*, 139–155. [CrossRef]

114. Sakshi Shashi Cerchione, R.; Bansal, H. Measuring the impact of sustainability policy and practices in tourism and hospitality industry. *Bus. Strategy Environ.* **2020**, *29*, 1109–1126. [CrossRef]

115. Han, H.; Yu, J.; Lee, J.S.; Kim, W. Impact of hotels' sustainability practices on guest attitudinal loyalty: Application of loyalty chain stages theory. *J. Hosp. Mark. Manag.* **2019**, *28*, 905–925. [CrossRef]

116. de Lange, D.; Dodds, R. Increasing sustainable tourism through social entrepreneurship. *Int. J. Contemp. Hosp. Manag.* **2017**, *29*, 1977–2002. [CrossRef]

117. Dyah Utami, D.; Dhewanto, W.; Dwi Lestari, Y.; Teknologi Bandung, I.; Aparicio, S. Rural tourism entrepreneurship success factors for sustainable tourism village: Evidence from Indonesia. *Cogent Bus. Manag.* **2023**, *10*, 2180845. [CrossRef]

118. Vecchio, R.; Annunziata, A.; Bouzdine-Chameeva, T. How to promote sustainable wine tourism: Insights from Italian and French young adults. *Ann. Tour. Res. Empir. Insights* **2024**, *5*, 100137. [CrossRef]

119. Martínez-Falcó, J.; Sánchez-García, E.; Marco-Lajara, B.; Millán-Tudela, L.A. Wine Tourism as a Catalyst for Sustainable Performance: The Mediating Role of Corporate Legitimacy and Green Innovation. *Agribusiness* **2024**, *early view*. [CrossRef]

120. Huang, S. A comprehensive science mapping of tourism and hospitality research: Tribes, territories and networks. *J. Hosp. Leis. Sport Tour. Educ.* **2025**, *36*, 100523. [CrossRef]

121. Sánchez-Franco, M.J.; Rey-Moreno, M. Do travelers' reviews depend on the destination? An analysis in coastal and urban peer-to-peer lodgings. *Psychol. Mark.* **2022**, *39*, 441–459. [CrossRef]

122. Harrington, R.J.; Ottenbacher, M.C. Strategic management: An analysis of its representation and focus in recent hospitality research. *Int. J. Contemp. Hosp. Manag.* **2011**, *23*, 439–462. [CrossRef]

123. Wang, C.; Xu, H.; Li, G. The corporate philanthropy and legitimacy strategy of tourism firms: A community perspective. *J. Sustain. Tour.* **2018**, *26*, 1124–1141. [CrossRef]

124. Ruiz-Ortega, M.J.; Córcoles-Muñoz, M.M.; Parra-Requena, G.; García-Villaverde, P.M. Sustainability orientation and sustainability performance in response to hostile environments in cultural tourism destinations. *Int. J. Tour. Cities* **2023**, *9*, 974–994.

125. Bilbao-Terol, A.; Arenas-Parra, M.; Cañal-Fernández, V.; Obam-Eyang, P.N. Multi-criteria analysis of the GRI sustainability reports: An application to Socially Responsible Investment. *J. Oper. Res. Soc.* **2018**, *69*, 1576–1598. [CrossRef]

126. Zhang, J. Impacts of the emissions policies on tourism: An important but neglected aspect of sustainable tourism. *J. Hosp. Tour. Manag.* **2021**, *47*, 453–461. [CrossRef]

127. Lin, M.S.; Zhang, H.; Luo, Y.; Li, Y. Environmental, social, and governance (ESG) measurement in the tourism and hospitality industry: Views from a developing country. *J. Travel Tour. Mark.* **2024**, *41*, 154–168. [CrossRef]

128. Lee, S.; Ham, S.; Yeon, J.; Lee, M. ESG themes of the restaurant industry: A comparison between full-service and limited-service restaurants. *J. Travel Tour. Mark.* **2024**, *41*, 20–34. [CrossRef]

129. Pérez Cañizares, P. "Corporate Sustainability" or "Corporate Social Responsibility"? A Comparative Study of Spanish and Latin American Companies' Websites. *Bus. Prof. Commun. Q.* **2021**, *84*, 361–385. [CrossRef]

130. Rhou, Y.; Singal, M. A review of the business case for CSR in the hospitality industry. *Int. J. Hosp. Manag.* **2020**, *84*, 102330. [CrossRef]

131. Kaur, P.; Talwar, S.; Madanaguli, A.; Srivastava, S.; Dhir, A. Corporate social responsibility (CSR) and hospitality sector: Charting new frontiers for restaurant businesses. *J. Bus. Res.* **2022**, *144*, 1234–1248. [CrossRef]

132. Yildiz, F.; Dayi, F.; Yucel, M.; Cilesiz, A. The Impact of ESG Criteria on Firm Value: A Strategic Analysis of the Airline Industry. *Sustainability* **2024**, *16*, 8300. [CrossRef]

133. Sim, M.; Kim, H. Measurement validation of a consumer-driven environmental, social, and governance (ESG) index for the airline industry. *J. Sustain. Tour.* **2024**, *33*, 474–499. [CrossRef]

134. Madanaguli, A.; Dhir, A.; Kaur, P.; Srivastava, S.; Singh, G. Environmental sustainability in restaurants. A systematic review and future research agenda on restaurant adoption of green practices. *Scand. J. Hosp. Tour.* **2022**, *22*, 303–330. [CrossRef]