

## Article

# The Contribution of Online Reviews for Quality Evaluation of Cultural Tourism Offers: The Experience of Italian Museums

Deborah Agostino <sup>1</sup>, Marco Brambilla <sup>2</sup>, Silvio Pavanetto <sup>3</sup> and Paola Riva <sup>1,\*</sup>

<sup>1</sup> Economics and Industrial Engineering, Department of Management, Politecnico di Milano, Via Lambruschini 4/b, 20156 Milano, Italy; deborah.agostino@polimi.it

<sup>2</sup> Data Science Lab, Informazione e Bioingegneria, Dipartimento di Elettronica, Politecnico di Milano, Via Ponzio 34/5, 20133 Milano, Italy; marco.brambilla@polimi.it

<sup>3</sup> Informazione e Bioingegneria, Dipartimento di Elettronica, Politecnico di Milano, Via Ponzio 34/5, 20133 Milano, Italy; silvio.pavanetto@polimi.it

\* Correspondence: paola.riva@polimi.it

**Abstract:** In the cultural tourism field, there has been an increasing interest in adopting data-driven approaches that are aimed at measuring the service quality dimensions through online reviews. To date, studies measuring quality dimensions in cultural tourism settings through content analysis of online user-generated reviews are mainly based on manual approaches. When the content analysis is automated, these studies do not compare different analytical approaches. Our paper enters this field by comparing two different automated content analysis approaches to evaluate which of the two is more adequate for assessing the quality dimensions through user-generated reviews in an empirical setting of 100 Italian museums. Specifically, we compare a ‘top-down’ content analysis approach that is based on a supervised classification built on policy makers’ guidelines and a ‘bottom-up’ approach that is based on an unsupervised topic model of the online words of reviewers. The resulting museum quality dimensions are compared, showing that the ‘bottom-up’ approach reveals additional quality dimensions compared with those obtained through the ‘top-down’ approach. The misalignment of the results of the ‘top-down’ and ‘bottom-up’ approaches to quality evaluation for museums enhances the critical discussion on the contribution that data analytics can offer to support decision making in cultural tourism.

**Keywords:** online user reviews; visitor perception; museum quality dimensions; user-driven quality dimensions; text modelling; online text analytics; user-generated content; data science; text mining; cultural tourism



**Citation:** Agostino, D.; Brambilla, M.; Pavanetto, S.; Riva, P. The Contribution of Online Reviews for Quality Evaluation of Cultural Tourism Offers: The Experience of Italian Museums. *Sustainability* **2021**, *13*, 13340. <https://doi.org/10.3390/su132313340>

Academic Editors: Sérgio Moro and Ricardo Filipe Ramos

Received: 20 October 2021

Accepted: 27 November 2021

Published: 2 December 2021

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In the cultural tourism field, there has been an increasing interest in adopting data-driven approaches to understand visitors’ perceptions (e.g., [1–7]). The research in this area offers several insights into the expectations of visitors [1], the opinions of travellers [8] or dimensions of service quality [9]. Although these analyses are quite diffused for touristic attractions such as hotels (e.g., [10]), there is much less evidence on the evaluation of quality dimensions as seen through users’ perceptions within museums. This is mainly because of the absence of a clear definition of the quality dimensions of museums [11] as opposed to what happens for hotels, where tourists’ perceptions are analysed with reference to predefined dimensions—such as cleanliness, location, room, value and service—usually displayed and specified by online review platforms (e.g., [10]).

Although still marginally studied, museums represent an important area of investigation for the tourism field because these institutions increase the attractiveness of a destination [12] and contribute to the economic development of touristic areas [13]. The literature on the identification of museum quality dimensions from online perceptions of museum visitors through online reviews is limited because most available contributions mainly focus on customer satisfaction analyses (e.g., [14,15]) and surveys (e.g., [16–18]).

Online reviews have long represented a valuable source for data analysis in the tourism field (e.g., [19,20]), but these data sources have been mostly studied in terms of the numerical ratings offered by review platforms in museum settings (e.g., [21]). Yet online reviews are mainly characterised by textual data, that is, comments written by visitors during their touristic experience. Although textual data represent valuable data sources for measuring the tourist's experience (e.g., [20]), the automated analysis of these data sources is scant within museum settings. Indeed, manual approaches to the analysis of online reviews have recently been used to investigate visitor perceptions, moving beyond customer satisfaction surveys (e.g., [9,22–24]). For example, the study by [9] explores the service quality dimension of museums from online reviews, but the content analysis is performed manually. Although automatic tools for text analytics have proven to be valuable in exploring quality dimensions in various applicative settings (e.g., [19,20]), to the best of our knowledge, there are still limited studies that analyse online reviews with the aim of automatically identifying quality dimensions of museums comparing the expectations of policy makers and the perceptions revealed by museum visitors through their own online voices.

To fill these gaps, the current study compares two different approaches to automated textual analysis of TripAdvisor data, here called the 'top-down' and 'bottom-up' approaches, with the aim of evaluating which one is more adequate for assessing quality dimensions through user-generated content in an empirical setting of 100 Italian museums. The 'top-down' approach is based on a predefined set of expected service quality dimensions, whereas the 'bottom-up' approach aims at identifying the latent dimensions of quality.

More specifically, the present study addresses the following RQs:

- RQ1: Which museum quality dimensions are identified following a 'top-down' approach for the analysis of online reviews?
- RQ2: Which museum quality dimensions are identified following a 'bottom-up' approach for the analysis of online reviews?
- RQ3: To what extent do the museum quality dimensions evaluated from online reviews using a 'bottom-up' approach differ from those identified through a 'top-down' approach?

The first research question (RQ1) evaluates the museum quality dimensions through a 'top-down' approach; this means that a predefined set of dimensions is defined by the decision maker (i.e., the policy maker), and we use a keyword-based classifier to analyse the expected dimensions in the text of online reviews. The second research question (RQ2) evaluates the museum quality dimensions through a 'bottom-up' approach; this means that latent quality dimensions have been directly derived from the textual description of the visitors' experiences by relying on Latent Dirichlet Allocation (LDA) [25], without imposing a predefined set of quality dimensions. The third research question (RQ3) compares the results of the two approaches, showing when it is more adequate to prefer a 'bottom-up' rather than a 'top-down' approach and, therefore, critically discussing the implications that different automated approaches to data analytics may have in supporting the decision making.

Our study contributes to the discussions on the impact that different data analytics approaches have in supporting organisations' decision making. The current paper is structured as follows: Section 2 presents the literature on the role of online user-generated data for quality assessment in the tourist field, with a specific focus on museums. Section 3 presents the methodology, detailing the available dataset and two analytical techniques adopted for analysing online data. Section 4 presents the results, which are critically commented upon in Section 5.

## 2. Literature Review

In the cultural tourism literature, online reviews have become a valuable data source for investigating the quality dimensions of the experience; these data offer the possibility of

collecting a huge amount of users' data without the need to ask visitors for this information, as these contents are voluntarily shared by visitors in a very personalised way [26]. This is in opposition to customer satisfaction surveys, which require the construction of questions and scales to evaluate dimensions of experiences through numerical ratings (e.g., [16–18,27–30]).

Alongside the recognition of the benefits of online reviews to understand users and their perceptions on cultural experiences, cultural tourism studies have grown significantly in recent years, and the literature on this topic can be divided into two main streams. The first stream exploits online reviews mainly using manual approaches, coding the content and manually classifying online reviews accordingly (e.g., [9,23]). The second stream exploits online reviews in an automatic manner, but the methodologies vary from one study to another. Some studies adopt 'top-down' approaches to online reviews' contents, searching automatically for predefined dimensions of quality in the dataset (e.g., [31–33]). Other studies adopt a 'bottom-up' approach to the content of online reviews searching for dimensions of the experience without defining an a priori set of quality dimensions (e.g., [7,24,26,34,35]).

The existence of different automated approaches to data analysis poses some questions on how they differ and whether one of the two approaches could be more appropriate than another [36]. Our study compares the 'top-down' with the 'bottom-up' approach to content analysis of online reviews, with the aim of understanding which of the two approaches is more adequate in exploring quality dimensions from online user-generated reviews.

The present study empirically applies to the context of museums, which are less investigated in the cultural tourism literature but represent an important area of investigation for the tourism field because they favour the cultural attractiveness of touristic destinations [12] and contribute to the economic development of touristic places [13].

### 3. Materials and Methods

This section describes the empirical context of the research (Section 3.1), the collection of online reviews (Section 3.2), a short description of the available dataset (Section 3.3) and the data analytics approaches to online reviews (Section 3.4).

#### 3.1. The Empirical Context of Italian Museums

The empirical context of the study is Italian state museums. The Italian context is particularly suited to ground cultural studies because UNESCO recognises Italy to be one of the countries with the highest density of cultural heritage sites [37]. Furthermore, in recent years, the Italian Ministry for Cultural Heritage and Activities and Tourism has been fostering the digital transformation of tourism as part of its strategic plan of development for 2017–2022 [38], pushing cultural institutions to develop digital strategies to promote cultural heritage and assets, monitor the dynamics of the brand reputation of cultural institutions and foster the diffusion of digital conversations connected to the culture.

In line with these directions, since 2018, the authors have been engaged in a project activated by the Italian Ministry for Cultural Heritage and Activities and Tourism, with the aim of monitoring the online reputation of a set of 100 Italian state museums that were selected by the ministry itself based on size, geographical distribution and type of collection exhibited. The authors have been engaged in the collection and analysis of museum qualities from online user-generated content. More specifically, the ministry identified a set of expected qualities of museums and asked the authors to verify the existence of these qualities within the online perception of the museums' public.

This allowed the authors to proceed with two parallel approaches to data analysis. On the one hand, the expected quality dimensions of museums defined by the policy maker have been searched within online reviewers' texts in a 'top-down' fashion, classifying online reviews in a supervised way according to the policy maker's expectations. On the other hand, the authors followed a 'bottom-up' approach to analyse in a data-driven and unsupervised fashion the text of online reviews; the aim was to identify the museum quality dimensions directly from the words of online users. The results of the 'top-down'

and ‘bottom-up’ analyses of the reviews presented within this paper allow the authors to discuss the potentialities of the ‘bottom-up’ approach in grasping the perceptions of quality dimensions directly from the users’ own words.

### 3.2. Data Collection

Data from online reviews have been collected from the TripAdvisor pages of the 100 Italian public museums that were selected by the Italian Ministry of Cultural Heritage and Activities and Tourism.

For each museum, we manually identified the TripAdvisor webpages and verified the credibility of the web sources directly with the museum managers. We then implemented the automated and scheduled data collection system, storing the online user-generated reviews in a document-based storage solution. This allowed the incremental update of the collections, enabling daily monitoring of the online reputation of museums. From the data collection system, we collect 47,993 online reviews published in 2019 on the TripAdvisor pages of the 100 Italian state museums.

Once collected, the online reviews were enriched through a language detection phase. The language of each online review was identified using a pre-trained Google model (implemented in the *Python* package *langdetect* [39]) with the help of an external service [40] to ensure the consistency of the results. The precision of the state-of-the-art language detection techniques is over 99% for 53 different languages (further details on the original project can be found at [41,42]).

To further validate the quality of the collected data and consistency of the data analytics, the results were also displayed in a dashboard. Granting real-time access to the dashboard, policy makers and museum managers could visualise, explore and monitor the online reputation of museums on TripAdvisor and on other online channels, such as online news websites and social media platforms such as Facebook, Instagram and Twitter. The real-time access to the dashboard also fostered frequent communication among policy makers, museum managers and researchers, thus allowing continuous quality validation of the analyses and results.

Overall, the data collection and enrichment procedure resulted in a dataset of 14,250 online reviews of museums automatically collected from TripAdvisor and for which the language has automatically been recognised to be Italian. This specific language has been selected not only because it was the most represented (30% of reviews) in the original dataset, but also to focus the research on local visitors of museums because the literature recognises the differences in the preferences of tourists because of their cultural background [11,43,44].

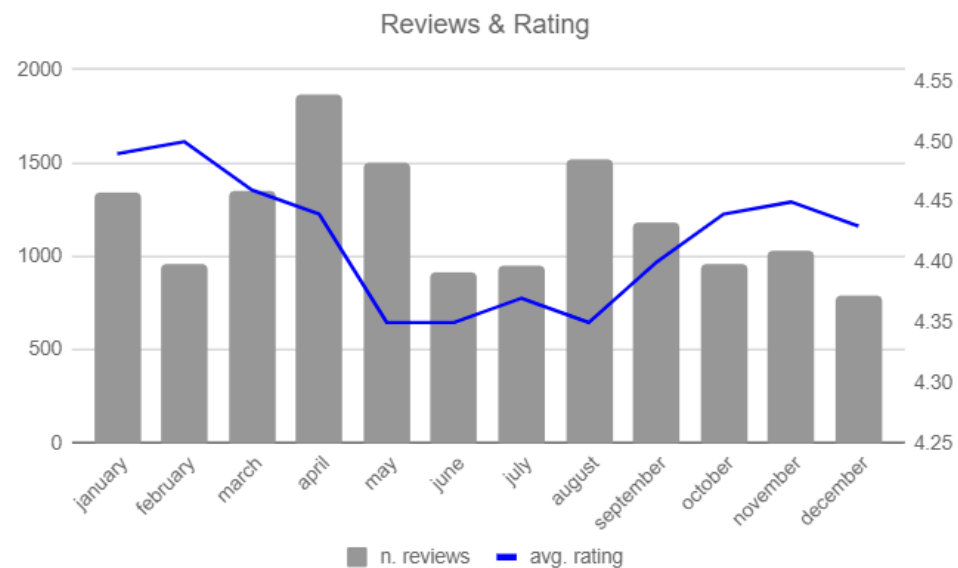
### 3.3. Online Reviews of Museums

The analysis of the Italian reviews shows seasonality in the amount of reviews and in the quantitative evaluation (i.e., rating) of the quality of the museum visits.

Looking at the review distribution over time (Figure 1), there is a peak in the spring and late summer, with 1854 reviews in April and 1506 in August. This can be connected to school trips and visits by foreign tourists, who tend to prefer spring and late summer to visit Italy [45]. On the contrary, the number of reviews decreases in winter, particularly in December (778 reviews) and February (953 reviews).

With reference to the quantitative value of the online ratings offered by TripAdvisor on a scale from 1 to 5, the results show a satisfactory evaluation of museums: the average rating of museums’ reviews over one year is 4.42 out of 5, with limited variability of the monthly average ratings because the values fall between 4.35 and 4.50 stars. However, the month-by-month distribution of the ratings shows values closer to 4.50 at the beginning of the year (January–March) than during late spring and summer, a period in which the average rating achieves minimum values (4.35 stars in August). Therefore, we observe a behaviour that is out of phase between the number of reviews and their ratings: periods like spring and late summer in which the number of reviews presents peaks, present low values in the average rating of reviews; the behaviour is reversed in autumn and winter, with few

reviews but highly rated. This could be connected to a more satisfactory perception of the museum experience when there are fewer people, hence in the quiet periods such as winter.



**Figure 1.** Evolution in the number of Italian reviews and average rating in 2019, monthly data.

### 3.4. Data Analytics Approaches to Online Reviews

In the current paper, there are two main approaches to analysing the text of online users' reviews: a 'top-down' approach and 'bottom-up' approach. Both approaches have been applied to the same dataset of 14,250 Italian reviews and are here thoughtfully described.

#### 3.4.1. 'Top-Down' Approach

The 'top-down' approach exploits online reviews by searching for predefined dimensions of the visit experience. It is called 'top-down' because we expect that these predefined dimensions are defined by the policy maker, who has some expectations on the quality of the service provided.

In our empirical context, the policy maker is represented by the Italian Ministry of Cultural Heritage and Activities and Tourism, which in 2018 introduced a set of quality standards for public museums with a Ministerial Decree (Ministerial Decree Nr. 113, 21 February 2018). These standards delineate a list of relevant aspects for which museums are held accountable (see policy makers' standards in Table 1). Based on this list and on the interviews with policy makers, we identified five quality dimensions that follow the 'top-down' perspective: *Ticketing and Welcoming*, *Space*, *Comfort*, *Activities*, and *Communication* (see 'Top-down' quality dimensions in Table 1). Each of these dimensions was associated with a set of keywords expected to be representative of that dimension. From an analytic perspective, the list of these keywords (see the set of keywords in Table 1) was used to build a keyword-based classifier for the classification of the text of online reviews into each of the five dimensions.

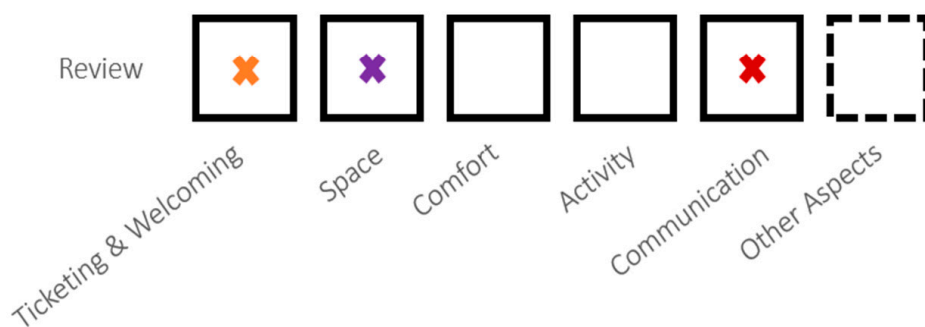
The implementation of the keyword-based classifier for the textual analysis can be interpreted as the automated version of the manual check performed by museum managers on the content of online reviews and exemplifies a 'top-down' approach. We built a non-overlapping multiclass keyword-based classifier to assign reviews to the five classes, that is *Ticketing and Welcoming*, *Space*, *Comfort*, *Activities* and *Communication*, based on the presence or absence of specific keywords in the text of the review (Table 1). Because the five 'top-down' quality dimensions are neither mutually exclusive nor exhaustive, one review can simultaneously be associated with more than one class or none (Figure 2). In the latter case, we label such a review with the term *Other Aspects* to underline that the review is not connected to any of the quality dimensions defined by the policy maker.

**Table 1.** ‘Top-down’ quality dimensions of museums as derived from the directions of the policy maker. Keywords are translated into English to increase readability and comprehension, but the algorithm uses the original words in Italian.

Policy Maker’ Standards	‘Top-Down’ Quality Dimension	Set of Keywords
Perception of museum user/visitor towards reception services, i.e., friendliness and professionalism of the staff, hostesses, stewards, security guards, cleaners, ticket office, presence of queues, queues, crowds, waiting, flow management, route management, etc. Perception of museum user/visitor towards costs of basic and additional services, i.e., costs incurred/to be incurred	Ticketing and Welcoming	Free, sliding, queue, free, ticket, throng, crowd, wait, entrance, steward, checkout, cost
Perception of museum user/visitor towards museum’s physical location Perception of museum user/visitor towards accessibility, e.g., lifts, platforms/slides for people with disabilities, transport to reach the place, parking lots	Space	Restoration, dirt, external/externally, intern, enter, unkempt, access, imposing, out
Perception of museum user/visitor towards use of spaces, e.g., halls, exhibitions, set-ups, lighting, signage, aesthetics, exhibition, captions, tour itinerary, cleaning, indoor locations, outdoor locations	Comfort	Lighting, degraded, cured, held
Perception of museum user/visitor towards organised activities and events, e.g., exhibitions, shows, presentations, experiences, workshops, educational activities, events, etc.	Activities	Guided tour, show, event
Perception of museum user/visitor towards usability, usefulness, completeness and quality of the applications and content of websites and social accounts, i.e., clarity, completeness and exhaustiveness of the information on the institution available online by consulting the institutional websites and social media/network accounts; quality and usability of the applications available online for booking and purchasing tickets for visiting institutes/exhibitions and the like; website compatibility with mobile devices	Communication	Guide, videoguide, audioguide, package, audio, video, booking, indication, itineraries

To select the classification algorithm for the text, we compared the keyword-based classifier with a Bidirectional Encoder Representations from Transformers (BERT) algorithm [46] specifically designed for the Italian language. We decided to use the language model BERT because this method recently caused a stir in the machine learning community, achieving state-of-the-art results in a wide variety of Natural Language Processing (NLP) tasks. Because we were specifically interested in the analysis of Italian reviews, we selected a BERT model pre-trained specifically on social media contents (i.e., Twitter messages) written in Italian [47]. Thanks to this choice, the model was already prepared for our empirical setting, avoiding further training of the model.

“The Miramare Castle is a beautiful castle rich in history immersed in a wonderful park and overlooking the cliff a few kilometers from Trieste. The park is easily **accessible**, rich in vegetation even if a little lacking as **indications**, the most scenic part is the one in front of the castle with the small dock and the beautiful fountain in the center of the square. I highly recommend a visit at sunset, the sunset gives a priceless spectacle! The position of the castle also offers a priceless view of the Gulf of Trieste. Paid parking (a bit expensive), some difficulties with the car in the busiest moments since you arrive from the **crowded** promenade.”



**Figure 2.** Example of the result of the application of the ‘top-down’ approach for the identification of the museum quality dimensions within online reviews, here based on a non-overlapping multiclass keyword-based classifier. The review is classified into three predefined museum quality dimensions out of the five dimensions defined by the policy maker.

To test the algorithmic performances of the keyword-based classifier and of the BERT language model, we randomly sampled 1000 Italian online reviews of museums and manually screened their text to assign a value of 1 to each ‘top-down’ category whenever the text was indeed addressing the aspects connected to the category. With stratified k-fold cross validation, we split the manually labelled data to obtain a training set of 800 reviews and a testing set of 200 reviews. In addition, thanks to the frequent monitoring of online reviews supported by the dashboard we developed (Section 3.2), we already expected highly unbalanced data, even before the application of the text classifier. This expectation was confirmed not only in the randomly sampled reviews used to test the performances, but also in the overall dataset considered for the analyses, as shown in the results (see also Section 4.2).

The performance of the algorithms are consequently affected by the highly unbalanced ‘top-down’ categories: the keyword-based method obtained an average accuracy of 80% and recall of 50% among the five classes (Table 2), while the BERT method obtained 88.2% accuracy and 58% recall. Notwithstanding the slightly higher performances of BERT, we selected the keyword-based classifier over the BERT method because the latter was greatly affected by the unbalanced nature of data being unable to predict three out of the five categories.

**Table 2.** Performance of keyword-based classifier for each of the ‘top-down’ quality dimensions and average across categories.

‘Top-Down’ Quality Dimension	Accuracy
Ticketing and Welcoming	72%
Space	76%
Comfort	80%
Activities	94%
Communication	75%
Average	80%

### 3.4.2. ‘Bottom-Up’ Approach

The ‘bottom-up’ approach exploits online reviews without a predefined expectation regarding the dimensions of the visit; rather, it is based on deriving the latent dimensions of the experience directly from the reviewers’ words without any predefined expectations.

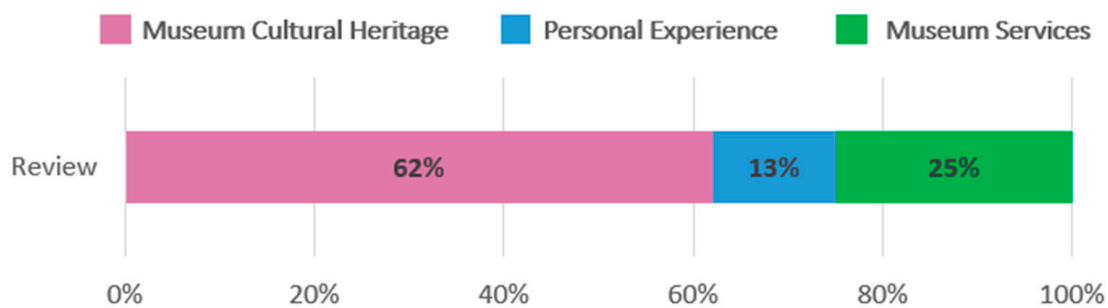
From an analytical perspective, the ‘bottom-up’ approach can be implemented through an unsupervised model based on topic modelling, here an LDA [25]. This model has been selected for two reasons. First, this generative probabilistic model entails the peculiar characteristics of Bayesian models of being highly flexible to the specific domain of the application. Second, this method allows us to detect hidden structures within the text of online reviews in terms of semantically similar groups of words, namely latent topics of discussion, that we interpret as latent museum quality dimensions that are hidden within the words of online reviewers. Indeed, by choosing the LDA procedure to define a set of topic-based quality dimensions, we are simulating the process of visitors in evaluating the quality of museums.

As far as the ‘bottom-up’ approach is concerned, we implemented an LDA in the *R* environment [48]. Resorting to the *tm* and *SnowballC* packages, the pre-processing phase consisted of converting text to lowercase, removing particular characters (e.g., emojis, URLs, punctuation and numbers), excluding language-specific and context-specific stopwords (i.e., roma, colosseo, pantheon, pantheum, phantheon, pompeii, firenze) and reducing the grammatical forms of the words through Porter’s stemming algorithm. Then, the four metrics proposed by [49–52] and implemented in the *FindTopicsNumber* function of the package *ldatuning* were used to select an appropriate number of topics between 2 and 30. Each of the plausible configurations of latent topics of discussion identified was interpreted by considering the 30 words with the highest probability values in the per-topic word distributions and reading the reviews with the highest probability values in the per-document topic distributions.

To increase the interpretability of the selected LDA model, we further grouped the resulting latent topics of the discussion into three main ‘bottom-up’ dimensions of museum quality, which we interpreted as *Museum Cultural Heritage*, *Personal Experience* and *Museum Services* (detailed description in Section 4.2). Thanks to the adoption of the LDA probabilistic topic model, the ‘bottom-up’ representation of each review is a mixture of three ‘bottom-up’ quality dimensions of museums, where the probability of observing a specific quality dimension is the emphasis of reviewers on the specific museum quality dimensions (Figure 3).



“The Miramare Castle is a **beautiful castle** rich in history immersed in a **wonderful park** and overlooking the cliff a few kilometers from Trieste. The park is easily accessible, rich in vegetation even if **a little lacking as indications**, the **most scenic part** is the one in front of the castle with the small dock and the **beautiful fountain** in the center of the square. I highly **recommend a visit** at sunset, the sunset gives a priceless spectacle! **The position of the castle** also offers a priceless view of the Gulf of Trieste. **Paid parking (a bit expensive)**, some difficulties with the car in the busiest moments since you arrive from the **crowded** promenade.”



**Figure 3.** Example of the result of the application of the ‘bottom-up’ approach for the identification of the museum quality dimensions within online reviews, here based on an LDA topic model. The review is a mixture of the three ‘bottom-up’ museum quality dimensions, where each proportion depends on the emphasis with which the reviewer discusses the corresponding quality dimension.

### 3.4.3. Comparison of the ‘Top-Down’ and ‘Bottom-Up’ Approaches

From a modelling perspective, the ‘top-down’ and ‘bottom-up’ approaches differ across at least five features: *Perspective*, *Categorisation*, *Training*, *Results interpretation* and *Representation* (Table 3).

**Table 3.** Comparison of the ‘top-down’ and ‘bottom-up’ approaches for evaluating museum quality dimensions.

	‘Top-Down’ Approach	‘Bottom-Up’ Approach
Perspective	Policy Maker/Museum Manager	Reviewer/User/Visitor
Categorisation	Supervised (keyword-based)	Unsupervised (topic-based)
Training	Required	Not required
Results interpretation	Not required	Required
Representation	Non-overlapping multiclass classifier output	Topic-based probability distribution

In terms of *Perspective* and *Categorisation*, the two approaches are complementary. The ‘top-down’ approach is a supervised model that simulates the behaviour of the policy maker, who defines a set of quality dimensions and desired to grasp how these are evaluated by the general public. Hence, the supervised model based on the specific set of keywords is just an automated version of the manual approach of reading and classifying data. On the contrary, the ‘bottom-up’ approach is an unsupervised topic-based model that simulates the museum visitor’s perspectives, identifying the quality dimensions of museums from

the latent dimensions of a museum visit detected within the words of online reviewers. This latter approach results in detecting a set of aspects not defined a priori and potentially new for the decision maker but that need to be interpreted. Because these aspects are hidden within the visitors' words, prior to the analysis, there is no clear indication of the number of dimensions or specific contents to be searched for. This is why a 'bottom-up' approach requires a certain effort for *Results interpretation*: once the analysis has been performed, it is necessary to interpret the resulting dimensions. On the contrary, once the keywords and the categories have been defined, the results of the 'top-down' approach can be immediately interpreted: each review is either associated or not with each of the specific dimensions, according to the identification of specific words within the text of the review.

From the point of view of *Training*, the two approaches have different analytical requirements. The 'top-down' approach needs to learn how to search for categories by starting from a training set of labelled reviews, whereas the 'bottom-up' approach learns the hidden structures directly from the data, without needing a training phase.

The two approaches also differ in terms of the output and, hence, of the *Representation* of each review. With the 'top-down' approach, each review is represented as a sequence of length given by the number of categories of the classifier, where each entry indicates in a binary way whether the specific 'top-down' dimension has been found or not in the text. With the 'bottom-up' approach, each review is still a sequence of length given by the number of dimensions retrieved, but each entry indicates the probability of referring to each specific dimension. Compared with the 'top-down' approach, the 'bottom-up' representation allows for each review to provide a ranking of the quality dimensions from the most to the least discussed to identify the most relevant and least relevant aspect discussed in each review. This allows for ranking reviews according to their propensity to discuss specific quality dimensions, while the 'top-down' approach is just able to detect whether a specific quality has been discussed or not.

#### 4. Results

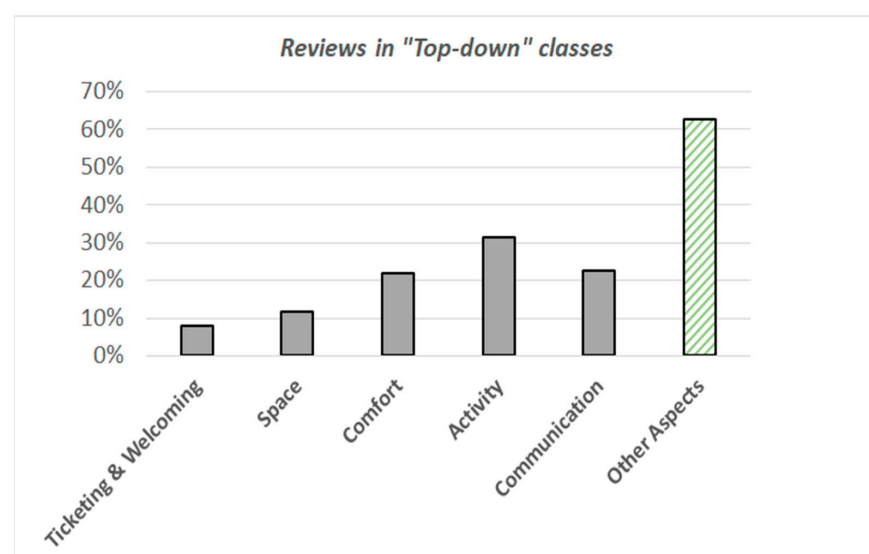
The results are presented in three main sections following the research questions. The first section (Section 4.1) presents the results of the 'top-down' approach to analyse online reviews, while the second section (Section 4.2) presents the quality dimensions obtained by adopting a 'bottom-up' approach. The third section (Section 4.3) critically discusses the (mis)alignment between the 'top-down' and 'bottom-up' museum quality dimensions.

##### 4.1. RQ1: Which Museum Quality Dimensions Are Identified following a 'Top-Down' Approach for the Analysis of Online Reviews?

The application of the 'top-down' approach resulted in a limited amount of reviews classified within the predefined five categories (Table 4), with 63% of the analysed reviews not assigned to any of the five museum quality dimensions identified by the policy maker and, therefore, labelled by us as belonging to the category *Other Aspects* (Figure 4). Manually scanning the content of these online reviews classified as *Other Aspects*, we found that these reviews were addressing many other aspects rather than the one identified by the policy maker: the five quality dimensions defined by the policy maker are related to the services offered by the museum, such as ticketing, communication and activities, while the museum public does not necessarily underline only these service-related aspects but rather refers to additional aspects.

**Table 4.** Short description of the ‘top-down’ classes and examples of excerpts of reviews classified in the corresponding class.

‘Top-Down’ Quality Dimension	Short Description	Excerpt of Review Classified
Ticketing and Welcoming	Aspects related to surveillance, welcoming, fees and costs, such as ticketing, queueing and crowding	‘... Paid parking (a bit expensive), some difficulties with the car in the busiest moments since you arrive from the crowded promenade.’
Space	Aspects related to the physical characteristics of the museum, such as the location of the museum and its accessibility	‘... beautiful castle rich in history immersed in a wonderful park and overlooking the cliff a few kilometres from Trieste. The park is easily accessible, rich in vegetation...’
Comfort	Aspects related to the equipment of museums’ exhibitions, such as lighting, cleanliness or maintenance	‘... the rooms with well-kept furnishings, paintings, furnishings that are well preserved and repaired from tampering ...’
Activities	Aspects related to events organised by museums, such as guided tours and temporary exhibitions	‘... The advice is to book a guided tour of at least 4 h, as we did, and you will not regret it, as a shorter time is really small...’
Communication	Aspects related to information offered to the public onsite or through online channels, such as physical signposts and audio-guides	‘... even if a little lacking as indications, the most scenic part is the one in front of the castle with the small dock and the beautiful fountain in the centre of the square...’

**Figure 4.** The proportion of Italian reviews that have been classified in each of the ‘top-down’ quality dimensions identified by policy makers (solid bars) and proportion of Italian reviews not associated with any ‘top-down’ quality dimension (striped bar). Notice that the percentages do not sum up to 100% because of the adoption of a non-overlapping multiclass classifier.

This result highlights that the ‘top-down’ approach supports the identification of specific quality dimensions of interest for the policy maker, here service-related dimensions, but fails in detecting the many other aspects of interest for museum reviewers: the interests of museum reviewers goes beyond the set of keywords predefined by the policy maker.

#### 4.2. RQ2: Which Museum Quality Dimensions Are Identified following a ‘Bottom-Up’ Approach for the Analysis of Online Reviews?

This section provides the results of the ‘bottom-up’ approach based on the application of an LDA model to the same dataset analysed in the previous section. Following this ‘bottom-up’ approach, we obtained 13 latent topics that we further interpreted as representing three ‘bottom-up’ quality dimensions (Table 5):

- *Museum Cultural Heritage (6 latent topics)*: With an average probability of 46%, the museum reviews address those aspects connected to the artistic collection of the museum, including comments on exhibitions, findings and artworks, but also considerations of the museums’ history and tradition and descriptions of the buildings, facades, churches and castles.
- *Personal Experience (4 latent topics)*: With an average probability of 31%, museum reviews address the emotional aspects associated with their personal experiences. This includes comments connected to the ‘wow effect’ of the visit, praises for the majesty and beauty and suggestions to visit the heritage site at least once in a lifetime. Additional aspects addressed are connected to the descriptions of revisits to the museum and the associated expectations, but also events that occurred during the visit or in connection to the visit itself, such as the museum’s disorganisation in supporting visitors or lack of information or encounters with rude personnel.
- *Museum Services (3 latent topics)*: With an average probability of 23%, the museum reviews address aspects connected to the services offered by museums, such as ticketing, guided tours, accessibility and transports.

**Table 5.** ‘Bottom-up’ quality dimensions derived from the Italian text of online museum reviewers. Italian stem words and reviews excerpts have been translated into English to increase readability and comprehension, but the algorithm elaborated on Italian texts.

‘Bottom-Up’ Quality Dimension	Most Probable Words (English Translation)	Latent Topics Names	Excerpt of Review
Museum Cultural Heritage	Museum, exposition, floor, room, collection, church, century, fresco, building, structure, marvel, wonder, beauty, artwork, art, gallery, masterpiece, castle seen, inner, landscape, panorama suggest, external, garden, villa, park, palace, fountain	Artistic Collection, Exhibitions and Findings, Castles and Views, Churches and Religious Antiquity, The Museum from Outside, Museum’s History and Tradition	<p>‘Castle built around 1850–60 in a medieval style wanted by the then Habsburg Empire, more precisely at the behest of Maximilian of Habsburg Archduke of Austria...’ (probability of observing <i>Museum Cultural Heritage</i> in review: 64%)</p> <p>‘... The museum presents a great variety of works ranging from painting, sculpture and architecture. In addition, you can admire some of the most famous masterpieces of the art world such as...’ (probability of observing <i>Museum Cultural Heritage</i> in review: 59%)</p> <p>‘... The temple was rebuilt in the form in which we can admire it today by the Emperor Hadrian (128 AD), under whose reign the Empire of Rome reached the height of its splendour...’ (probability of observing <i>Museum Cultural Heritage</i> in review: 69%)</p>

Table 5. Cont.

'Bottom-Up' Quality Dimension	Most Probable Words (English Translation)	Latent Topics Names	Excerpt of Review
Personal Experience	Visit, day, suggest, time, place, emotion, experience, see, years, remain, pity, just, personal, tourist, unfortunately	Emotional Visits, Revisits and Expectations, At least once!, Unfortunately	<p>'... And every time, looking up at its internal vault, I am amazed at how a closed place can convey that sense of immense space and deep breath to me' (probability of observing <i>Personal Experience</i> in review: 44%)</p> <p>'I remembered seeing a marvel in a state of abandonment, this year I revisited it and I was amazed...' (probability of observing <i>Personal Experience</i> in review: 41%)</p> <p>'... Reviewing the excavations is a great thrill every time, despite the poor signage and maps that are not always clear. However, the organisation turned out to be even worse...' (probability of observing <i>Personal Experience</i> in review: 47%)</p>
Museum Services	See, guide, appreciate, organise, path, accompany, explain, tour, ticket, entry, euro, queue, reservation, cost, price, entry, site, park, reach, close, foot, walk, convenient, easy	Accessibility and Transports, Guided Tours, Ticketing (purchase, price, book)	<p>'... You can have free and privileged access to the cash desks... the site is easily accessible from Naples by bus...' (probability of observing <i>Museum Services</i> in review: 46%)</p> <p>'... Our guide was extremely engaging, able to actualise what we saw...' (probability of observing <i>Museum Services</i> in review: 40%)</p> <p>'... Having arrived 1 h earlier, we got in line anyway! Once the tickets have been taken, we are told the entrance...' (probability of observing <i>Museum Services</i> in review: 51%)</p>

The identification of these three 'bottom-up' dimensions from the museum reviews shows that the museum visitors emphasise various aspects of the experience beyond the services identified by the policy makers. Specifically, the 'bottom-up' analysis reveals that the museum reviewers consider cultural heritage aspects and personal experiences when evaluating the quality of the museum experience (Table 6): on average, an online review of museums discusses more about museum cultural heritage aspects (46% average probability) and personal experiences (31% average probability) rather than museum services (23% average probability).

These results are relevant for both policy makers and museum experts because the 'bottom-up' approach reveals the necessity to consider not only service-related aspects here such as the 'top-down' service dimensions, but also cultural heritage and personal experiences, which naturally emerge from the 'bottom-up' approach towards the analysis of museum reviews. The misalignment between the 'bottom-up' and 'top-down' results already prepares the way for a discussion of the bias that museum experts and policy makers may introduce in evaluating museum quality dimensions using a 'top-down' approach, which is a focal aspect of the following section.

**Table 6.** Summary statistics of the probability distribution of the three ‘bottom-up’ dimensions of museum quality as obtained from the sum of the per-document topic probability distributions of each topic associated with the corresponding ‘bottom-up’ dimension.

‘Bottom-Up’ Quality Dimension	Min	Q1	Q2	Mean	Q3	Max
Museum Cultural Heritage	15.31%	41.83%	46.15%	46.12%	50.12%	83.54%
Personal Experience	9.37%	27.31%	30.55%	30.80%	33.94%	58.46%
Museum Services	6.47%	19.59%	22.05%	23.09%	25.61%	59.46%

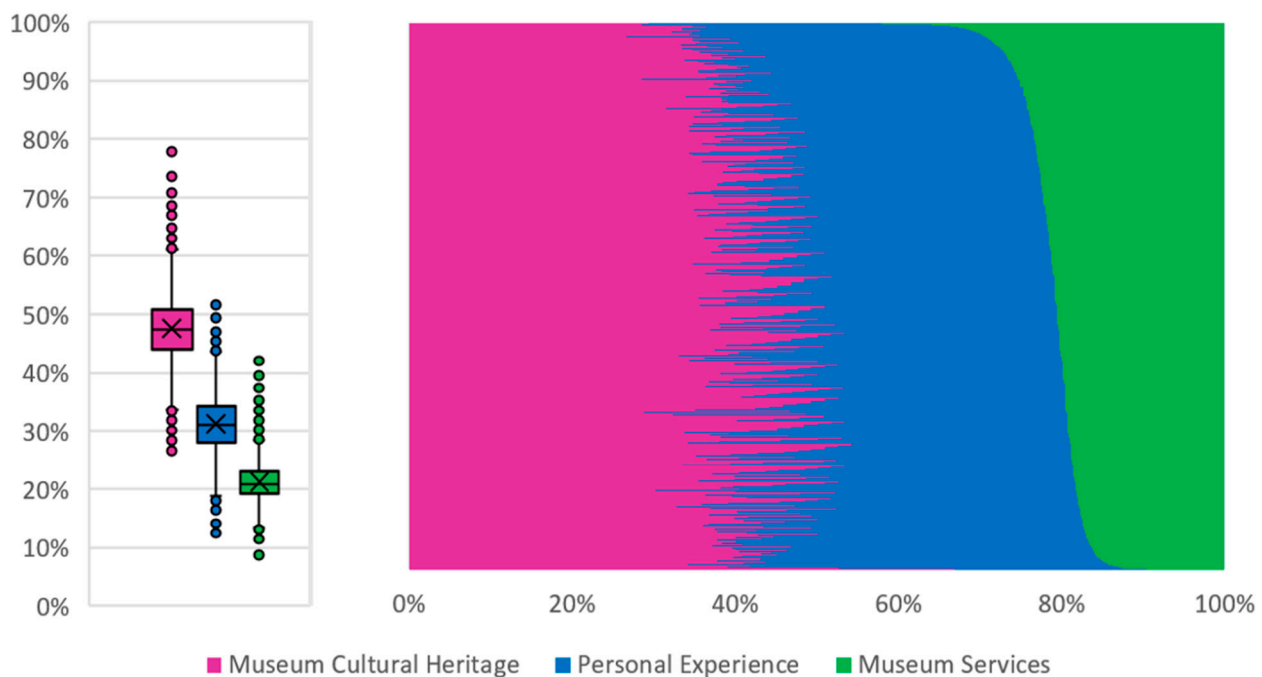
#### 4.3. RQ3: To What Extent Do the Museum Quality Dimensions Evaluated from Online Reviews Using a ‘Bottom-Up’ Approach Differ from Those Identified through a ‘Top-Down’ Approach?

The ‘top-down’ and ‘bottom-up’ approaches show different results, both in terms of the implementation of the method and results obtained. As far as the implementation of the methodology is concerned, the ‘top-down’ approach is based on a set of keywords all connected to museum services, which are defined from the standards issued by the policy maker; this approach resulted in 63% of online reviews that did not fit into any of the predefined quality dimensions (*Other Aspects*). The ‘bottom-up’ approach overcomes this limitation by searching for the aspects of interest using reviewers’ own words, without even acknowledging how many or which could be the quality dimensions of a museum: the quality dimensions of museums perceived by the reviewers are grasped as those aspects on which reviewers pose more emphasis when describing their experiences through their own words. These hidden perspectives are captured through an LDA and show that, on average, a museum review discusses more about a museum’s cultural heritage aspects (46% average probability) and personal experiences (31% average probability) than the services offered by the museum (23% average probability).

To further understand the differences between the ‘top-down’ and ‘bottom-up’ approaches, we focus on the reviews classified in a ‘top-down’ fashion as *Other Aspects* and look at the ‘bottom-up’ museum quality dimensions these reviews present (Figure 5). Using the ‘top-down’ approach to analyse these reviews, the policy maker would not have been able to detect any of the service quality dimensions of museums or grasp the aspects of actual interest to museum visitors. Using a ‘bottom-up’ approach, the policy makers would be able to explore the hidden aspects discussed by the online reviewers of museums without any predefined categories. From the empirical analysis, the most discussed aspects by the museum reviewers are connected to the heritage of the museum (48% average probability of observing *Museum Cultural Heritage*) and the personal experiences felt during the visit (31% average probability of observing *Personal Experience*), while attention to museum services is limited (21% average probability of observing *Museum Services*). Going along with the case of museums’ reviews classified as *Other Aspects*, the ‘bottom-up’ analysis reveals a high probability of addressing the latent aspects connected to the museum’s history (8.5% average probability of observing latent topic *Museum’s History and Tradition*) and to artworks (8.1% average probability of observing the latent topic *Artistic Collection*), but it also frequently refers to the emotions felt during the cultural experience (8.5% average probability of observing latent topic *Emotional Visits*).

It is important to note that the ‘bottom-up’ approach does not exclude the possibility of identifying service-related aspects if they are aspects of interest for the reviewers. Considering the case of museums’ reviews classified as *Other Aspects*, the ‘bottom-up’ analysis recognises *Museum Services* with a 21% average probability of observing the dimension within reviews. This means that in this specific case, the policy maker would also be able to detect the aspects related to services through the ‘bottom-up’ approach. Moreover, the analysis of the latent aspects connected to this dimension reveals an average probability of observing the latent topic *Accessibility and Transports* equal to 7.4%, *Guided Tours* to 7.2% and *Ticketing (purchase, price, book)* to 6.6%. Below are examples (translated in English for clarity) of Italian reviews classified through the ‘top-down’ approach as addressing

*Other Aspects* but that show a high probability of discussing the ‘bottom-up’ dimension of *Museum Services*.



**Figure 5.** Comparison of the distributions of the ‘bottom-up’ museum quality dimensions over reviews classified as addressing *Other Aspects* when using the ‘top-down’ approach.

*From the Porta San Paolo railway station (from the Piramide metro stop), take the train to Ostia Antica, after a journey of about half an hour. In ancient times, it was the ancient port of Rome, and in it, the goods flowed and passed to and from the whole empire. The ruins are well preserved, and all the activities of the time are recognisable from them, from the storage warehouses to the bathrooms. public buildings, amphitheatres, fire stations and port corporations. I leave the rest to your curiosity. I bet you will be charmed. (30.7% probability of observing the latent topic Accessibility and Transports)*

*Nice initiative by the students of the Rodolico scientific high school. We were welcomed with kindness and cordiality by the students, appreciating their competence. (20.1% probability of observing latent topic Guided Tours)*

*Admission 9 euros per person and 4 euros for children seems a bit excessive to me .. to possibly add 1 euro for transport by bus because if you proceed on foot, the path to take is not at all simple and in good shape. Strollers are impossible!!! (19.2% probability of observing latent topic Ticketing (purchase, price, book))*

## 5. Discussion and Conclusions

The current paper has compared two different approaches to the automated textual analysis of online user-generated reviews, here called ‘top-down’ and ‘bottom-up’, with the aim of evaluating which of the two is more adequate for the assessment of quality dimensions through user-generated contents, empirically setting the research on the 14,250 TripAdvisor Italian reviews received by 100 Italian state museums in 2019.

The ‘top-down’ approach is based on a predefined set of expected service quality dimensions that are defined by the decision maker (i.e., the policy maker); once defined, these dimensions are automatically searched for within the dataset of online reviews, here by implementing an automated supervised keyword-based non-overlapping multiclass classifier for the Italian text of the reviews.

The 'bottom-up' approach identifies the latent quality dimensions emerging from the visitors' own words; this is implemented by modelling the text through an unsupervised topic model, namely an LDA [25], applied to the online words of the reviewers. This means that latent quality dimensions have been directly derived from the textual description of the visitors' experiences, without imposing a predefined set of quality dimensions.

Comparing the two approaches, differences emerge in terms of both the implementation of the methodology and the results obtained from the empirical analysis of the Italian museum reviews.

As far as implementation is concerned, the two approaches differ in terms of the *Perspective, Categorisation, Results interpretation, Training and Representation* of each review.

The 'top-down' approach is categorised as supervised because it requires a specific set of quality dimensions to be defined a priori; because these dimensions are predefined by the policy maker, this approach offers the decision maker a focused perspective when identifying these quality dimensions. Once the quality dimensions have been defined, the 'top-down' approach requires training the model for the automated classification of the text of the reviews into these predefined dimensions. Only after a training phase is the model able to automatically assign a review to a specific dimension: when keywords are found in the text of the review, the review is recognised as referring to that dimension. Using this binary method to represent whether a review refers or not to a specific dimension, the results of the 'top-down' approach are of immediate interpretation for the decision maker: a review automatically assigned to specific dimensions will refer to those dimensions of interest for the decision maker.

The 'bottom-up' approach is categorised as unsupervised because it identifies quality dimensions by modelling text without requiring any a priori definition of the dimensions. Therefore, without requiring a training phase, the 'bottom-up' approach automatically learns the quality dimensions by recognising the latent structures within the text of the reviews. Because these quality dimensions are automatically detected from the online reviewers, they capture the latent perspective of the users. Because of this user perspective, the decision maker is required to put some effort into interpreting the latent quality dimensions, but once these dimensions have been interpreted, each review is represented through the probability of referring to each of the interpreted dimensions. This representation allows the decision maker to rank the quality dimensions from the most to the least emphasised and identify the most and least emphasised aspect for each review; this implies that decision makers can identify which quality dimensions are perceived as the most relevant by users and which reviews most emphasise the specific dimensions that could be of interest to control.

As far as the empirical application of the two approaches to online reviews is concerned, the approaches offer different insights.

The 'top-down' approach identifies the occurrence of the five service quality dimensions that are predefined by the policy maker (i.e., *Ticketing and Welcoming, Space, Comfort, Activities and Communication*) within the text of the reviews. However, the results show that 63% of the reviews did not refer to any of the predefined quality dimensions because they were classified as discussing *Other Aspects*. This finding presents a potential risk for policy makers adopting the 'top-down' approach when analysing the reviews because the interest of museum reviewers goes beyond the set of keywords predefined by the policy maker.

Instead, the 'bottom-up' approach identifies 13 latent dimensions that have been interpreted as defining the three main quality dimensions, called *Museum Cultural Heritage, Personal Experience* and *Museum Services*. We have also observed an average predominance of emotional and heritage aspects of the visit experience compared with the services provided by museums. This finding underlines that according to the visitors' perspectives, the museums' quality dimensions are not only limited to museum services, but they also include those aspects connected to cultural heritage assets and personal experiences felt during the visit, which, on average, are more relevant than museum services in the evaluation of the experience at museums.



### 5.1. Academic Implications

From an academic perspective, the current paper provides two main implications.

First, the present paper enhances the debate on the contribution of data analytics to tourism management (e.g., Rita et al., 2018), showing that an automated approach to data analysis matters: comparing two different approaches to online user-generated data, we find that several differences exist, not only in terms of the implementation phases required, but most importantly, in terms of the results obtained. This finding has relevant implications for data-driven decision making because it suggests that the decision maker should be aware of the approach through which the users' data are analysed to reduce the information bias connected to the analytical procedure used to analyse the data. This is shown finding that those aspects considered as quality dimensions by the decision maker can be highly different from those aspects perceived as quality dimensions by final users: using a 'top-down' approach within the specific setting of museums, most of the reviews (63%) do not relate to the museum service quality dimensions defined by the policy maker because museum visitors cherish quality dimensions beyond just those of museum services (23%), placing more emphasis instead on cultural heritage (46%) and personal experiences (31%).

The second implication relates to the cultural tourism literature, with particular reference to the debate around the identification of quality dimensions for museums (e.g., [16]). Although most tourism studies investigate and assess the quality dimension of touristic attractions as hotels (e.g., [10]), our paper focuses on the less studied but touristic relevant setting of museums, highlighting the existence of different quality perspectives. In the museum context, we show that users' perspectives include the services offered by the museum, such as ticketing and the communication of internal activities, as well as the experience offered to the visitor, for example, visiting the museum more than once, and the characteristics of the heritage assets, such as the collection exposed or the museum's building. This finding suggests that evaluating the quality dimensions of museums based only on the services offered represents a limitation in the museum context because personal experiences and heritage assets are perceived as relevant dimensions by museum visitors. Nonetheless this finding, our study does not aim at providing a punctual list of quality dimensions for museums: considering the personal narratives of users' experiences, we have been able to identify three main museum quality dimensions, but we also recognise that quality dimensions can be emergent and differ depending on the user who is performing the review. Specifically, our empirical application showed 13 latent dimensions, but another investigation on different users or time periods on the same heritage sites could potentially produce other quality dimensions.

### 5.2. Practitioner Implications

Our study also offers two major implications for practitioners.

The first implication relates to the existence of the different *implementations* required for the adoption of each approach, either 'top-down' or 'bottom-up'. This difference significantly influences the choices of policy makers and museum managers who are in charge of exploiting online user-generated data to identify and assess service quality. Our study provides practical guidance on the implementation of 'top-down' or 'bottom-up' approaches by detailing the differences between these two methodologies in terms of their perspectives, requirements, training, representation of outputs and interpretations. These practical aspects can support policy makers and museum managers who are interested in applying this methodology to exploit online user-generated data. Notwithstanding the methodology adopted, it is important to underline that professional knowledge of data analytics competences is required to analyse online data. This poses some challenges on the professional profiles inside museums, which typically include architects, archaeologists, managers and registrars but less often individuals with analytical competences.

The second implication refers to the existence of different *results* from one approach to another in terms of the identification of quality dimensions. Here, the same dataset can

result in different quality dimensions depending on the automated analytical approach. When commissioning these studies or analyses, both policy makers and museum managers should be aware of the type of approach adopted because this can provide different results and differently support decision making. We are not arguing that one of the two approaches is better than the other, but we are saying that depending on the purpose of the analysis, one method can be better suited than the other. If the intent is to search for some quality dimensions to understand how many visitors perceived some specific aspects, such as service-related aspects, then a ‘top-down’ approach should be preferred because it selects just those reviews explicitly connected with the few aspects fixed by the decision maker. Instead, if the intent is to understand which aspects are relevant for visitors to evaluate quality, a ‘bottom-up’ approach should be preferred because it provides quality dimensions as hidden structures among the words of online users and does so without any a priori assumption. This latter approach may be helpful in rapidly evolving situations, such as the current COVID-19 pandemic: policy makers willing to understand the new dimensions of quality perceived by users could use a ‘bottom-up’ approach to automatically derive them from their own words.

### 5.3. Limitations and Further Research

The current study has two major limitations, which, if properly addressed, may lead to future developments. First, the current study focuses only on Italian reviews, limiting the generalisability of the results to only local visitors to Italian museums. Although the Italian language was the most represented in the original dataset (30% of the reviews in Italian), Italian museums present reviews in more than 30 languages. Extensions of this work could analyse the differences in the quality dimensions of museums across language groups to study the behaviour of nonlocal visitors of museums, who are claimed to be potentially different than local visitors [44]. A second limitation of the current study is associated with the set of museums analysed, which is represented just by Italian state museums. Future research could investigate the validity of our studies for museums in other countries, within specific nations or across national borders, or consider museums with other governance forms, such as foundations or corporate museums.

**Author Contributions:** Conceptualisation, D.A. and P.R.; data curation, S.P. and P.R.; formal analysis, P.R.; funding acquisition, D.A. and M.B.; investigation, P.R.; methodology, M.B. and P.R.; project administration, D.A. and P.R.; resources, M.B.; software, S.P. and P.R.; supervision, D.A. and P.R.; validation, D.A., M.B., S.P. and P.R.; visualisation, P.R.; writing—original draft preparation, D.A., M.B. and P.R.; writing—review and editing, D.A. and P.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors would like to thank the Italian Ministry of Cultural Heritage and Activities and Tourism for the collaboration shown along the design of the supervised classification described in this study.

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

## References

1. Sheng, C.-W.; Chen, M.-C. A study of experience expectations of museum visitors. *Tour. Manag.* **2012**, *33*, 53–60. [[CrossRef](#)]
2. Schuckert, M.; Liu, X.; Law, C.H.R. Hospitality and Tourism Online Reviews: Recent Trends and Future Directions. *J. Travel Tour. Mark.* **2015**, *32*, 608–621. [[CrossRef](#)]
3. Bi, J.-W.; Liu, Y.; Fan, Z.-P.; Zhang, J. Wisdom of crowds: Conducting importance-performance analysis (IPA) through online reviews. *Tour. Manag.* **2018**, *70*, 460–478. [[CrossRef](#)]

4. Jia, S. Motivation and satisfaction of Chinese and U.S. tourists in restaurants: A cross-cultural text mining of online reviews. *Tour. Manag.* **2020**, *78*, 104071. [[CrossRef](#)]
5. Antonio, N.; Correia, M.; Ribeiro, F. Exploring User-Generated Content for Improving Destination Knowledge: The Case of Two World Heritage Cities. *Sustainability* **2020**, *12*, 9654. [[CrossRef](#)]
6. Kirilenko, A.P.; Stepchenkova, S.O.; Dai, X. Automated topic modeling of tourist reviews: Does the Anna Karenina principle apply? *Tour. Manag.* **2020**, *83*, 104241. [[CrossRef](#)]
7. Taecharungroj, V.; Mathayomchan, B. Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. *Tour. Manag.* **2019**, *75*, 550–568. [[CrossRef](#)]
8. Hou, Z.; Cui, F.; Meng, Y.; Lian, T.; Yu, C. Opinion mining from online travel reviews: A comparative analysis of Chinese major OTAs using semantic association analysis. *Tour. Manag.* **2019**, *74*, 276–289. [[CrossRef](#)]
9. Su, Y.; Teng, W. Contemplating museums' service failure: Extracting the service quality dimensions of museums from negative on-line reviews. *Tour. Manag.* **2018**, *69*, 214–222. [[CrossRef](#)]
10. Liu, Y.; Teichert, T.; Rossi, M.; Li, H.; Hu, F. Big data for big insights: Investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews. *Tour. Manag.* **2017**, *59*, 554–563. [[CrossRef](#)]
11. Falk, J.H.; Dierking, L.D. *The Museum Experience Revisited*, 1st ed.; Routledge: New York, NY, USA, 2016; pp. 1–416. [[CrossRef](#)]
12. Giglio, S.; Bertacchini, F.; Bilotta, E.; Pantano, P. Using social media to identify tourism attractiveness in six Italian cities. *Tour. Manag.* **2019**, *72*, 306–312. [[CrossRef](#)]
13. Torre, A.; Scarborough, H. Reconsidering the estimation of the economic impact of cultural tourism. *Tour. Manag.* **2017**, *59*, 621–629. [[CrossRef](#)]
14. Eleni, M.; Costantine, L. Factors affecting museum visitors' satisfaction: The case of greek museums. *Tourismos* **2013**, *8*, 271–287.
15. Moreno Gil, S.; Brent-Ritchie, J.R.; Almeida-Santana, A. Museum tourism in Canary Islands: Assessing image perception of Directors and Visitors. *Mus. Manag. Curatorship* **2019**, *34*, 501–520. [[CrossRef](#)]
16. Nowacki, M.; Kruczek, Z. Experience marketing at Polish museums and visitor attractions: The co-creation of visitor experiences, emotions and satisfaction. *Mus. Manag. Curatorship* **2020**, *36*, 62–81. [[CrossRef](#)]
17. Oren, G.; Shani, A.; Poria, Y. Dialectical emotions in a dark heritage site: A study at the Auschwitz Death Camp. *Tour. Manag.* **2020**, *82*, 104194. [[CrossRef](#)]
18. Richards, G.; King, B.; Yeung, E.Y.M. Experiencing culture in attractions, events and tour settings. *Tour. Manag.* **2020**, *79*, 104104. [[CrossRef](#)]
19. Bi, J.-W.; Liu, Y.; Fan, Z.-P.; Zhang, J. Exploring asymmetric effects of attribute performance on customer satisfaction in the hotel industry. *Tour. Manag.* **2019**, *77*, 104006. [[CrossRef](#)]
20. Galati, F.; Galati, R. Cross-country analysis of perception and emphasis of hotel attributes. *Tour. Manag.* **2019**, *74*, 24–42.
21. Fernández-Hernández, R.; Vacas-Guerrero, T.; García-Muiña, F.E. Online reputation and user engagement as strategic resources of museums. *Mus. Manag. Curatorship* **2020**, 1–16. [[CrossRef](#)]
22. Ferguson, M.; Pichè, J.; Walby, K. Bridging or fostering social distance? An analysis of penal spectator comments on Canadian penal history museums. *Crime Media Cult. Int. J.* **2015**, *11*, 357–374. [[CrossRef](#)]
23. Wight, A.C. Visitor perceptions of European Holocaust Heritage: A social media analysis. *Tour. Manag.* **2020**, *81*, 104142. [[CrossRef](#)]
24. Simeon, M.I.; Buonincontri, P.; Cinquegrani, F.; Martone, A. Exploring tourists' cultural experiences in Naples through online reviews. *J. Hosp. Tour. Technol.* **2017**, *8*, 220–238. [[CrossRef](#)]
25. Blei, D. Probabilistic topic models. *Commun. ACM* **2012**, *55*, 77–84. [[CrossRef](#)]
26. Guo, Y.; Barnes, S.; Jia, Q. Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent Dirichlet allocation. *Tour. Manag.* **2017**, *59*, 467–483. [[CrossRef](#)]
27. Reyes-García, M.E.; Criado-García, F.; Camúñez-Ruiz, J.A.; Casado-Pérez, M. Accessibility to cultural tourism: The case of the major museums in the city of Seville. *Sustainability* **2021**, *13*, 3432. [[CrossRef](#)]
28. Xu, Z.; Zhang, H.; Zhang, C.; Xu, M.; Dong, N. Exploring the Role of Emotion in the Relationship between Museum Image and Tourists' Behavioral Intention: The Case of Three Museums in Xi'an. *Sustainability* **2019**, *11*, 559. [[CrossRef](#)]
29. Villaespesa, E. Museum Collections and Online Users: Development of a Segmentation Model for the Metropolitan Museum of Art. *Visit. Stud.* **2019**, *22*, 233–252. [[CrossRef](#)]
30. Frochot, I.; Hughes, H. HISTOQUAL: The development of a historic houses assessment scale. *Tour. Manag.* **2000**, *21*, 157–167. [[CrossRef](#)]
31. Borghi, M.; Mariani, M.M. Service robots in online reviews: Online robotic discourse. *Ann. Tour. Res.* **2020**, *87*, 103036. [[CrossRef](#)]
32. Tsai, C.-F.; Chen, K.; Hu, Y.-H.; Chen, W.-K. Improving text summarization of online hotel reviews with review helpfulness and sentiment. *Tour. Manag.* **2020**, *80*, 104122. [[CrossRef](#)]
33. Mehraliyev, F.; Kirilenko, A.P.; Choi, Y. From measurement scale to sentiment scale: Examining the effect of sensory experiences on online review rating behavior. *Tour. Manag.* **2020**, *79*, 104096. [[CrossRef](#)]
34. Shao, J.; Ying, Q.; Shu, S.; Morrison, A.M.; Booth, E. Museum Tourism 2.0: Experiences and Satisfaction with Shopping at the National Gallery in London. *Sustainability* **2019**, *11*, 7108. [[CrossRef](#)]
35. Dimache, A.; Wondirad, A.N.; Agyeiwaah, E. One museum, two stories: Place identity at the Hong Kong Museum of History. *Tour. Manag.* **2017**, *63*, 287–301. [[CrossRef](#)]

36. Rita, P.; Rita, N.; Oliveira, C. Data science for hospitality and tourism. *Worldw. Hosp. Tour. Themes* **2018**, *10*, 717–725. [CrossRef]
37. UNESCO. Available online: <https://whc.unesco.org/en/list/&order=country#alphaI> (accessed on 22 November 2021).
38. MIBACT. Available online: <https://www.turismo.beniculturali.it/en/home-strategic-plan-for-tourism/> (accessed on 20 February 2021).
39. Langdetect. Available online: <https://github.com/Mimino666/langdetect> (accessed on 22 November 2021).
40. Dandelion API. Available online: <https://dandelion.eu/docs/api/datatxt/li/> (accessed on 22 November 2021).
41. Google Language Detection. Available online: <https://code.google.com/archive/p/language-detection/> (accessed on 22 November 2021).
42. Google Language Model. Available online: <https://www.slideshare.net/shuyo/language-detection-library-for-java> (accessed on 22 November 2021).
43. Imamoğlu, Ç.; Yılmazsoy, A.C. Gender and locality-related differences in circulation behavior in a museum setting. *Mus. Manag. Curatorship* **2009**, *24*, 123–138. [CrossRef]
44. Trinh, T.T.; Ryan, C. Visitors to Heritage Sites. *J. Travel Res.* **2016**, *56*, 67–80. [CrossRef]
45. Guizzardi, A.; Mazzocchi, M. Tourism demand for Italy and the business cycle. *Tour. Manag.* **2010**, *31*, 367–377. [CrossRef]
46. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv Prepr.* **2018**, arXiv:1810.04805.
47. BERT Italian Pre-Trained. Available online: <https://huggingface.co/neuraly/bert-base-italian-cased-sentiment> (accessed on 22 November 2021).
48. Harrell, F.E. rms: Regression Modelling Strategies. R Package Version 5.1-3.1. 2019. Available online: <https://CRAN.R-project.org/package=rms> (accessed on 3 September 2021).
49. Griffiths, T.L.; Steyvers, M. Finding scientific topics. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 5228–5235. [CrossRef]
50. Cao, J.; Xia, T.; Li, J.; Zhang, Y.; Tang, S. A density-based method for adaptive LDA model selection. *Neurocomputing* **2009**, *72*, 1775–1781. [CrossRef]
51. Arun, R.; Suresh, V.; Veni Madhavan, C.E.; Narasimha Murthy, M.N. On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations. In *Advances in Knowledge Discovery and Data Mining*; Zaki, M.J., Yu, J.X., Ravindran, B., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; Volume 6118. [CrossRef]
52. Deveaud, R.; Sanjuan, E.; Bellot, P. Accurate and effective latent concept modeling for ad hoc information retrieval. *Doc. Numérique* **2014**, *17*, 61–84. [CrossRef]