The impact of Pampulha Modern Ensemble UNESCO nomination as a world cultural heritage on TripAdvisor's reviews

Rafael Almeida de Oliveira[a], Renata Maria Abrantes Baracho[a], & Alexandre Miguel de Andrade Souza [b]

[a]School of Information Science
UFMG – Universidade Federal de Minas Gerais, Brazil
rafalolbh@hotmail.com

[b]Secretaria de Estado de Desenvolvimento Social de Minas Gerais, Brazil

alexandre.andrade@social.mg.gov.br

Abstract

The study aims to evaluate the impact of Pampulha Modern Ensemble (Belo Horizonte, Brazil) after its nomination as a world cultural heritage, from the users' reviews and comments on the TripAdvisor website. The information collected from the creation of a web scraper enabled the extraction of data to a structured database and subsequent analysis with the use of text mining techniques. The results showed that the nomination has had a direct effect on the user satisfaction, especially in the first five months after the designation. Despite the positive data, it was noticed that users still do not have a consolidated image of all the attractions of the building complex as a heritage of humanity, as it lacks strategies that propagate them as integrated products, adding value to the sites.
Keywords: big data analytics; text analytics; web scraping; heritage; pampulha.

1 Introduction

Since the 1990s, there has been a considerable increase in the volume of data available from different communication channels, especially the Internet (Miltner & Gray, 2014). This large volume, in addition to storage speed, availability and varied data formats, was named after the phenomenon known as big data (Mcafee & Brynjolfsson, 2012; Miltner & Gray, 2014). For data to become information that may aid people and organisations in decision-making, some basic tools are needed such as servers and ways of management and computational techniques that help information organisation, analysis and visualisation (Andrejevic, 2014; Chen, Chiang, & Storey, 2012).

It is a fact that most big data is availably spread all over the Internet by users of social networks, especially from their comments and reviews of products and services (Oliveira, 2017). In the travel industry, for instance, TripAdvisor is the most significant worldwide travel community that brings meaningful content related to tourism and is the most popular platform for tourist information, mainly because it encourages its members to leave comments and write reviews of destinations and tourist services (Xiang, Schwartz, Gerdes & Uysal, 2015; Yoo, Sigala, & Gretzel, 2016). User opinions can help business owners and managers understand, quantitatively and qualitatively, what customers are saying about their establishment and destinations (Torres, 2013).

Web scrapers or scraping are tools used to extract online data from TripAdvisor and other websites, to be organised and analysed eventually. They convert unstructured website data, created in programming languages, into structured local databases (Vargiu & Urru, 2012). After structuring the bank, it is possible to analyse user ratings and comments by using techniques that enable them to extract value from the data and assist in decision-making. Text analytics or text mining are techniques used to analyse comments on social networks (Gandomi & Haider, 2015), generating quantitative data from texts. They also help find textual patterns in unstructured data sources, enabling their understanding more rationally (Moreno & Redondo, 2016).

Among the attractions registered on TripAdvisor, there are a few cultural heritage sites of humanity. Since 1972, the United Nations Educational, Scientific and Cultural Organization (UNESCO) has nominated cultural heritage that is essential for understanding the history of humanity, to ensure the protection, conservation and preservation of culture for future generations (Labadi, 2017). In Brazil, Pampulha Modern Ensemble was named by UNESCO on July 17, 2016. The ensemble, created in the city of Belo Horizonte in the 1940s, is a landmark of Brazilian modernism and inspiration for various works in the world, bringing together buildings and gardens linked to leisure and culture. In addition to an artificial lake, the complex has a former casino (transformed into the current Pampulha Art Museum), the Yacht Club, the Ballroom and St. Francis of Assisi Church, created by architect Oscar Niemeyer with the help of several other artists such as Portinari and Burle Marx (UNESCO, 2017).

As part of the modern ensemble, Pampulha Lake was first built as a dam to provide water to the population in 1938. After the election of the mayor Juscelino Kubitschek (JK) in the 1940s, who was fond of modernist ideas for urban development, he invited the architect Oscar Niemeyer to design projects for the church, the ballroom, the Yacht Club and the casino. Their objective was to convert the old dam into an artificial lake with major works of art on its surroundings and make it a fresh and clean social space for leisure (Araujo, 2015). Therefore, the urbanist architecture became the symbol of Pampulha and after the election of Juscelino Kubitschek as the president of Brazil, Pampulha's model was replicated again by Oscar Niemeyer during the construction of Brasilia, the new capital of Brazil, in the early 1960s (Resende, 2004).

Monitoring heritage information is crucial for the elaboration of strategies that allow the efficient monitoring of the development of these spaces (Chianese, Piccialli, & Valente, 2015). From this view, this article aims to examine if there is any impact on Pampulha Modern Ensemble as a World Heritage Site after its nomination, based on the collection and analysis of reviews and comments from TripAdvisor users, between 2015 and 2018. The information collected is expected to enrich discussions about visitors' and residents' perceptions from a UNESCO nomination, generating inputs for monitoring similar attraction data and future research on the subject.

2 Literature Review

2.1 Big Data analytics

The popularisation of digital devices such as smartphones and sensors resulted in an extensive production of data, hitherto non-existent in society, increasing the need for tools that enable its interpretation in real-time (Gandomi & Haider, 2015). Therefore, the enormous potential of big data is to transform data into information that enables insights for decision-making (Gandomi & Haider, 2015) and generate new knowledge (Khan & Vorley, 2017).

The process of extracting big data insights from automated tools is known as big data analytics (Gandomi & Haider, 2015). Unstructured data such as text, images, audios and videos make up 95% of big data (Gandomi & Haider, 2015) and it is a major challenge to analyse these sources efficiently. This data has the role of complementing information,

collected traditionally, such as surveys and archival data (Xiang et al., 2015).

Gandomi and Raider (2015) list five primary kinds of data analysis of big data:

Text analytics (text mining): Techniques for extracting information from digitally present textual data fed by users of websites or organisations. The text analysis involves statistical analysis, computational linguistics, and machine learning.

Audio analytics: Extract and analyse information from unstructured audio data. It is performed by transcribing the audio into text or by the phonetic-based approach.

Video analytics: A variety of techniques for extracting data, analysing and monitoring video information from smart algorithms, metadata, transcripts and visual content. Audio and text analytics techniques may be applied.

Social media analytics: Analysis of data from social media channels, including review sites such as TripAdvisor. It works with two types of information: user-generated content (feelings, images, videos) and relationships between network entities (people, organisations, and products).

Predictive analytics: techniques that predict future outcomes based on historical and current data using statistical analysis.

Among the types of the analysis presented, text analytics stands out for works related to content generation by social networks, such as TripAdvisor.

2.2 Text analytics

Much of the unstructured information on the Internet, such as e-mails, blogs, forums, and especially social networks, comes from texts generated by users (Moreno & Redondo, 2016). These texts contain information ranging from comments on service and product delivery to free-topic reports on specific subjects. Therefore, the analysis of this data through text analytics plays a key role in big data analytics (Xiang et al., 2015).

Text analytics tools help to identify relevant information in large databases, previously not effectively visible for knowledge creation in

organisations (Khan & Vorley, 2017). They are used in many industries because the data analysed helps decision-makers understand market dynamics and fraud risks (Moreno & Redondo, 2016), as well as facilitate knowledge management strategies (Khan & Vorley, 2017).

Gandomi and Raider (2015) highlight four text analytics methods:

Information extraction: techniques for the extraction of structured data from unstructured text such as scraping;

Text summarisation: production of one or more texts summaries to highlight higher terms, words and subjects of the content analysed;

Question answering: provides answers to questions asked in natural language;

Feelings analysis: contain people's opinions on products, services, organisations, individuals and events. *Document-level techniques* determine whether a document expresses a feeling that is negative or positive. *Sentence-level techniques* attempt to determine the polarity of thoughts about an entity expressed in a single sentence. *Aspect-based techniques*, which are widely used in product reviews, identify the feeling within a document, classify and evaluate it, so it can be measured.

Moreno and Redondo (2016) add five other methods, in addition to information extraction, summarization and question answering techniques:

Topic Tracking and Detection: From documents read by the user, it identifies keywords of new documents related to the texts, suggesting similar content to the reader;

Categorisation or classification: Identifies the main topics of the document or the most repeated words in the text. Requires prior knowledge to classify terms;

Clustering: it is a technique used to put together different documents in a way that they are merged without using predefined topics or categories. A cluster allows a document to appear under several subtopics. This way, a document is not omitted from a search result, for instance.

Concept linkage: Lists documents identifying common concepts among them and helps users find information that would not be possible in traditional search methods;

Information visualisation: places a large amount of text hierarchically visualised or word map, enabling to view topics of more or less relevance to the document. A *tag cloud* is a well-known example of information visualisation;

Deep learning: A combination of Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN) techniques that assist in predicting sequential information, such as which word should appear in a sequence from words already written in the same sentence;

Text analytics is a continually evolving area, fundamental for big data analysis. In other words, it is a door for understanding content created by millions of users in social networks (Moreno & Redondo, 2016), leading this generation of information to assume various types of business.

2.3 Web Scraping

Digital data extraction has become one of the most widely used techniques in big data (Marres & Weltevrede, 2013). Considering that individual pages of a website have the same information in the same templates, a web scraper can understand the logic of positioning this information, collecting it from different pages and passing the information to a unique and analysable database (Marres & Weltevrede, 2013).

The great advantage of the technique is that in addition to creating large databases with countless variables, it also creates, from countless types of information, more synthetic and manageable databases (Landers, Brusso, Cavanaugh, & Collmus, 2016). Thus, scraping tools expand the sociological potential of the Internet, once information from large social networks can be analysed (Marres & Weltevrede, 2013). Without these tools, a student or researcher would take some considerable time to navigate through all pages manually and even review the material and insert it into a data, almost an impossible task without collection automation (Landers et al., 2016).

In general, the scraping process is composed of a series of steps in which formatted data is extracted from a large amount of unorganised information. It also gives details on parts of a text or the way these texts are arranged for future analysis in an available database (Marres & Weltevrede, 2013).

Despite the increased use of scraping from the government, the private sector and academia do not have laws that address the issue directly, creating a grey area of debate (Krotov & Silva, 2018). It is common for most non-governmental websites to include in their terms of use a ban on the use of tools such as robots and web scrapers to collect information by their users to avoid violating their copyrights (Krotov & Silva, 2018). However, a site does not necessarily own the data created by its users, such as ratings and comments, which are widely used in social networks (Krotov & Silva, 2018). Besides, user ideas cannot be copyrighted; only the specific forms of representation of these ideas are guaranteed rights and allow fair use of this information on a limited scale (Krotov & Silva, 2018).

Therefore, the use of data from scrapers should be done collectively and not individually, by analysing patterns of behaviour of a society or group, from the context in which individuals are inserted. From this view, individuality will not be compromised, ensuring the ethics of research (Andrejevic, 2014).

3 **Methodology**

The survey was conducted using web scrapers that enabled information extraction (Gandomi & Haider, 2015) available on the following Belo Horizonte's attractions on TripAdvisor: Pampulha Lake, Pampulha Modern Ensemble, St. Francis of Assisi Church, Ballroom and Yacht Club. Only Yacht Club did not provide relevant results for the analysis as it is a private leisure club, only visited by its members. Therefore, it has no regular visitors for tourism purposes, which may explain the low number of reviews and comments on this attraction. Initially, we used a web scraper available online, without programming, called *import.io*. However, over time, the previously free web scraper placed limitations on the project, creating the need to develop its web scraper from the python programming language. This same web scraper not only made it possible to continue extracting user reviews as in the explanation above, but also allowed user comments to be extracted for each attraction individually.

Figure 1 shows the fields that were extracted for all 49 attractions of Belo Horizonte, including those related to Pampulha Modern Ensemble, allowing comparison of results between the complex and the other attractions of the city. Figure 1 presents the information collected

regarding user ratings (red line) and user comments, with the headings (blue lines) and the comment text box (yellow lines).

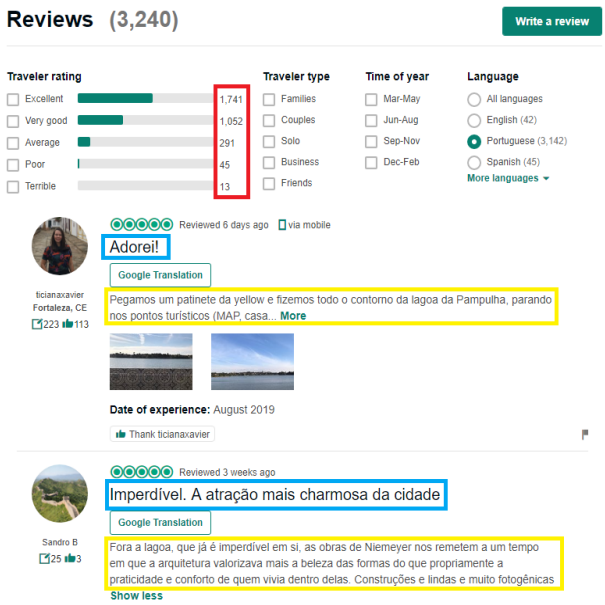

Fig 1. Information extracted from TripAdvisor

Extractions for the collection of user evaluation were performed in different stages. The first was held on July 12, 2016, five days before UNESCO's announcement. The second extraction was performed approximately 20 days after the announcement, and lastly, extractions were made on December 2016, July 2017, December 2018 and July 2018, a total of two-year time after the announcement.

The comments and titles were extracted in May 2019. Since each comment shows the date it was posted on, it was possible to classify them by year, differently from the users' evaluation. The information collected in the comment text box did not detect any impact from the collection of keywords related to the naming of Pampulha's attractions. However, the results of the words contained in the comment headings were relevant to the analysis. Thus, the analysis of the titles was performed from a text summarization (Gandomi & Haider, 2015), adding words that were repeated between titles, demonstrating them from a descending ranking, according to the information visualisation method (Gandomi & Haider, 2015). It is noteworthy that the analysis of the texts was performed by users who evaluated the attractions between grades 4 and 5, since the

number of evaluations between 1 and 3 was deficient. Finally, only the texts published on the platform in the Portuguese language were analysed.

4 **Results**

Data extraction, after five days of the announcement of Pampulha as a cultural heritage of humanity, showed that three attractions of the complex stood out among the ten attractions of the city with the highest growth in the number of reviews: Pampulha Lake, St. Francis of Assisi Church and Ballroom (Table 1). On average, each attraction in Belo Horizonte increased its value by 5.95%. The attraction Pampulha Lake had the most significant variation of all the attractions of the city reaching an increase of 14.36% in the number of reviews. Even after five months of the announcement, Pampulha Lake remained the most positive popular attraction. The attraction, despite losing a position at the end of the first year of analysis and another position at the end of the second year, still had very positive results compared to the average popularity of Belo Horizonte. As for the three other attractions highlighted in the August's extraction, they appeared among the ten best-evaluated attractions during the time series, but showing inconsistency (Table 1). It is noteworthy that Pampulha Modern Ensemble did not appear in the evaluation results. There is a hypothesis that, as this area is popularly known only as Pampulha Lake, the specific attraction of Pampulha Modern Ensemble is less searched on TripAdvisor for reviews.

Regarding the analysis of the comments, it was possible to extract all texts from each Belo Horizonte's attraction. Thus, unlike the number of assessments and ratings of attractions extracted from the value fields on the main page of each attraction, the analysis of aggregated data was made by year, as the comments showed the date they had been posted since the attractions were included on the website. As the objective of the research is to analyse the impact of Pampulha's results from the 2016 appointment, the analysis of the comments started in 2015.

In the case of Pampulha Modern Ensemble, it is clear that the word 'heritage' appeared as a highlight in the commentary titles in 2016, 2017 and 2018 (Table 3). As early as 2015, before the appointment of UNESCO, the word does not appear.

The same results are found in Pampulha Modern Ensemble for St. Francis of Assisi Church. The word 'heritage', which was not prominent in 2015, has appeared steadily since 2016, among the users who gave the attraction ratings 4 and 5 (Table 4). The results of the keywords for Ballroom, Pampulha Art Museum and Pampulha Lake did not highlight terms related to the heritage in none of the years analysed.

Table 1. Percentage change of assessments of the main attractions in Belo Horizonte

Attraction	Aug 16/ Jul 16	Attraction	Dec 16/ Jul 16
Belo Horizonte	**5,95**	**Belo Horizonte**	**21,50**
Pampulha Lake	**14,36**	**Pampulha Lake**	**57,67**
Topazio Theater	10,60	Kubitschek House	42,48
Maletta Building	9,91	Topazio Theater	38,41
Guanabara Park	9,55	Maletta Building	32,53
Niemeyer Building	9,34	Natural Sciences Museum	31,16
Israel Pinheiro Square	9,24	Diamond Mall	30,63
Natural Sciences Museum	8,70	Minascentro	30,09
Belo Horizonte Central Market	8,01	Belo Horizonte Central Market	29,96
St. Francis of Assisi Church	**6,85**	**Ballroom**	**28,57**
Pampulha Modern Ensemble	**6,83**	Guanabara Park	28,28

Attraction	Jul 17/ Jul 16	Attraction	Jul 18/ Jul 17
Belo Horizonte	**37,02**	**Belo Horizonte**	**11,79**
Diamond Mall	104,97	Diamond Mall	28,22
Pampulha Lake	**96,54**	Kubitschek House	23,81
Kubitschek House	67,26	**Pampulha Lake**	**21,15**
Natural Sciences Museum	62,32	Belo Horizonte Central Market	18,75
Belo Horizonte Central Market	56,17	Banco do Brasil Cultural Center	18,71
Topazio Theater	52,98	**Pampulha Art Museum**	**16,90**
Banco do Brasil Cultural Center	47,54	Natural Sciences Museum	16,52
Israel Pinheiro Square	44,54	Minas Gerais Vale Memorial	16,27
Minascentro	43,32	Niemeyer Building	15,59
Ballroom	**43,29**	**St. Francis of Assisi Church**	**13,53**

Source: TripAdvisor. Created by the authors

Table 2. Percentage change of ratings of the main attractions in Belo Horizonte

Attraction	Aug 16/ Jul 16	Attraction	Dec 16/ Jul 16
Belo Horizonte	**0,03**	**Belo Horizonte**	**0,17**
Ballroom	**0,60**	**Ballroom**	**1,46**
Guanabara Park	0,46	Administrative City	1,13
Raul Soares Square	0,44	**Pampulha Lake**	**0,90**
Pampulha Modern Ensemble	**0,37**	**Pampulha Modern Ensemble**	**0,74**
Pampulha Lake	**0,37**	Guanabara Park	0,73
Topazio Theater	0,35	Banco do Brasil Cultural Center	0,62
Pampulha Art Museum	**0,35**	**Pampulha Art Museum**	**0,61**
Minascentro	0,26	Kubitschek House	0,57
St. Francis of Assisi Church	**0,23**	Maletta Building	0,49
UFMG Knowledge Space	0,21	**St. Francis of Assisi Church**	**0,48**

Attraction	Jul 17/ Jul 16	Attraction	Jul 18/ Jul 17
Belo Horizonte	**0,20**	**Belo Horizonte**	**-0,06**
Ballroom	**1,89**	**Ballroom**	**0,61**
Diamond Mall	1,38	Diamond Mall	0,54
Guanabara Park	0,97	Luiz de Bessa Public Library	0,53
Administrative City	0,94	UFMG Knowledge Space	0,50
Banco do Brasil Cultural Center	0,88	Belo Horizonte Central Market	0,37
Pampulha Modern Ensemble	**0,79**	Administrative City	0,30
Independence Arena	0,78	Minas Gerais Vale Memorial	0,24
Church of São José	0,77	Church of São José	0,22
Belo Horizonte Central Market	0,77	Niemeyer Building	0,21
Museum of Arts & Offices	0,62	Guanabara Park	0,19

Source: TripAdvisor. Created by the authors

Table 3. Top primary words: Pampulha Modern Ensemble (ratings 4 and 5)

2015		2016		2017		2018	
Word	**N**	**Word**	**N**	**Word**	**N**	**Word**	**N**
pretty	115	pretty	109	pretty	42	pretty	22
worth it	35	architecture	59	architecture	26	place	14
architecture	30	**heritage**	**47**	**heritage**	**15**	architecture	12
unmissable	25	humanity	27	worth it	13	beautiful	8
postcard	18	beautiful	25	unmissable	10	unmissable	6
architectural	18	unmissable	23	beautiful	7	architectural	5
art	14	art	21	architectural	7	Niemeyer	4
beautiful	11	postcard	16	humanity	6	**heritage**	**4**

Source: TripAdvisor. Created by the authors

Table 4. Top primary words: St. Francis of Assisi Church (ratings 4 and 5)

2015		2016		2017		2018	
Word	**N**	**Word**	**N**	**Word**	**N**	**Word**	**N**
pretty	97	pretty	174	pretty	53	beautiful	24
architecture	40	architecture	78	art	35	pampulha	8
little church	31	work	65	work	21	work	7
art	29	art	54	unmissable	17	little church	5
card	29	**heritage**	**42**	**heritage**	**16**	unmissable	4
postcard	29	beautiful	32	little church	16	architecture	4
beautiful	26	little church	23	postcard	15	postcard	3
work	24	postcard	21	visit	7	marriage	3
unmissable	12	unmissable	17	architecture	4	wonderful	3
faith	12	modern	11	beautiful	4	visit	3
worth it	10	Portinari	10	worth it	4	**heritage**	**2**

Source: TripAdvisor. Created by the authors

5 Discussion

Pampulha Modern Ensemble's appointment as a cultural heritage of humanity has had a positive impact on TripAdvisor's users, especially in

the short term. This impact was most noticeable in the level of user satisfaction. Interestingly, in the space between a month and five months after the announcement by UNESCO, there was no infrastructure investment or significant improvements in the region that could increase user satisfaction. It may be suggested that only the announcement of the complex as a World Heritage Site itself has caused visitors to appreciate the region's attractions.

In the specific case of the number of reviews, it is believed that the appointment will bring positive results to Pampulha Lake. It is because the attraction, besides giving its name for the entire region, comprises not only the modern heritage site but also a zoo, a hiking trail, an ecological park, among others. Thus, it can be explained that the nomination helped to increase the region's assessments as shown, however, once the comments are analysed qualitatively, they did not highlight a direct relationship with the word ‘heritage’, but only strengthened terms for the landscaping and scenic beauty of the lake and its surroundings.

The specific analysis of Pampulha Modern Ensemble clearly showed the strong relationship of the world heritage in the title of user comments from 2016 on. That is, there was, in fact, recognition of the site as heritage by users. This data was also observed for St. Francis of Assisi Church, considered the most significant tourist symbol of the complex, helping to identify it as a world heritage site.

However, it is noted that users’ perception of UNESCO’s nomination for other attractions is still limited (Ballroom and Pampulha Art Museum). Therefore, it suggests that there is a more significant promotion by the city of these attractions as an integral part of the complex considered as heritage. Thus, there would be a higher chance of increasing the number of visitors to these spaces (increasing the number of reviews on TripAdvisor), as well as generating a better interpretation of the monuments as a unique travel itinerary. This will facilitate their appreciation and preservation as a major attraction for the city, benefiting the products and services of the region and generating greater competitiveness to the destination.

6 **Conclusion**

This study aimed to evaluate the impact of the naming of Pampulha Modern Ensemble as a World Heritage Site by UNESCO from user

reviews and comments on TripAdvisor. From web scraping techniques for data extraction and further analysis by text mining and information visualisation methods, it was concluded that up to the first five months of the appointment, user satisfaction results were positive for all the attractions of the complex. There was also a higher perception of visitors in the commentary titles regarding the word 'heritage' for the main attractions of the complex. However, it suggests that campaigns to promote the other attractions should be carried out so that visitors recognise them more clearly as belonging to the complex considered as World Heritage, increasing visitation and preservation of spaces, and creating a stronger identity among these attractions.

This work leaves as a future possibility the application of techniques for comparing results between different attractions considered as heritage, besides increasing the collection of information from other travel sites considered as user-commenting platforms, such as Expedia or Google.

Financial Support

This study was partly sponsored by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (Capes) - Finance Code 001.

7 References

1. Andrejevic, M. (2014). The big data divide. *International Journal of Communication*, *8*(1), 1673–1689.
2. Araújo, W. M. (2015). Os "turistas moradores" no Complexo Arquitetônico da Pampulha em Belo Horizonte: experiência mediada pelo design e a arquitetura. *RITUR-Revista Iberoamericana de Turismo, 5,* 18-28.
3. Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *Mis Quarterly*, *36*(4), 1165–1188. https://doi.org/10.1145/2463676.2463712
4. Chianese, A., Piccialli, F., & Valente, I. (2015). Smart environments and Cultural Heritage: a novel approach to create intelligent cultural spaces. *Journal of Location-Based Services*, *9*(3), 209–234. https://doi.org/10.1080/17489725.2015.1099752
5. Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*. https://doi.org/10.1016/j.ijinfomgt.2014.10.007
6. Khan, Z., & Vorley, T. (2017). Big data text analytics: an enabler of knowledge management. *Journal of Knowledge Management*, *21*(1), 18–34. https://doi.org/10.1108/JKM-06-2015-0238
7. Krotov, V., & Silva, L. (2018). Legality and ethics of web scraping. In *Twenty-fourth Americas Conference on Information Systems*. New Orleans.
8. Labadi, S. (2017). UNESCO, World Heritage, and Sustainable Development: International Discourses and Local Impacts. In P. G. Gould & K. A. Pyburn (Eds.), *Collision or Collaboration* (pp. 45–60). Springer. https://doi.org/10.1007/978-3-319-44515-1

9. Landers, R. N., Brusso, R. C., Cavanaugh, K. J., & Collmus, A. B. (2016). A Primer on Theory-Driven Web Scraping: Automatic Extraction of Big Data from the Internet for Use in Psychological Research Richard. *Psychological Methods*, *21*(4), 475–492. https://doi.org/10.1037/a0033269
10. Marres, N., & Weltevrede, E. (2013). Scraping the social? *Journal of Cultural Economy*, *6*(3), 313–335. https://doi.org/10.1080/17530350.2013.772070
11. Mcafee, A., & Brynjolfsson, E. (2012). Big Data: The Management Revolution, 2012. *Harvard Business Review*, (October), 1–9.
12. Miltner, K., & Gray, M. L. (2014). Crawford K, Miltner K and Gray M (2014) Critiquing Big Data politics, ethics, epistemology.pdf. *International Journal of Communication*, *8*, 1663–1672.
13. Moreno, A., & Redondo, T. (2016). Text Analytics: the convergence of Big Data and Artificial Intelligence. *International Journal of Interactive Multimedia and Artificial Intelligence*, *3*(6), 57. https://doi.org/10.9781/ijimai.2016.369
14. Oliveira, R. (2017). *Extração de dados web como suporte na elaboração de indicadores do turismo de Minas Gerais: uma iniciativa em Big Data*. Universidade Federal de Minas Gerais, Brazil.
15. Resende, M. S. A. D. (2004). *O conjunto da Pampulha em Belo Horizonte: concepção e usos para o lazer e turismo* (1943/2003).
16. Torres, J. L. X. de S. (2013). *Analisis cuantitativo de los hoteles en Tripadvisor : Destinos costeros en España y Portugal*. Universidad de Malaga.
17. UNESCO. (2017). Pampulha Modern Ensemble. Retrieved September 2nd, 2019 from https://whc.unesco.org/en/list/1493/
18. Vargiu, E., & Urru, M. (2012). Exploiting web scraping in collaborative filtering-based approach to web advertising. *Artificial Intelligence Research*, *2*(1), 44–54. https://doi.org/10.5430/air.v2n1p44
19. Xiang, Z., Schwartz, Z., Gerdes, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*. https://doi.org/10.1016/j.ijhm.2014.10.013
20. Yoo, K.-H., Sigala, M., & Gretzel, U. (2016). *Exploring TripAdvisor*. (R. Egger, I. Gula, & D. Walcher, Eds.), *Springer*. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-54089-9