

Repositório ISCTE-IUL

Deposited in *Repositório ISCTE-IUL*:

2022-05-19

Deposited version:

Accepted Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Silva, B., Moro, S. & Marques, C. (2021). Sensing the impact of COVID-19 restrictions from online reviews: The cases of London and Paris unveiled through text mining. In Reis, J. L., Parra López, E., Moutinho, L., and Santos, J. P. M. dos. (Ed.), *Marketing and Smart Technologies. Smart Innovation, Systems and Technologies*. (pp. 223-232). La Laguna: Springer Singapore.

Further information on publisher's website:

[10.1007/978-981-16-9268-0_18](https://doi.org/10.1007/978-981-16-9268-0_18)

Publisher's copyright statement:

This is the peer reviewed version of the following article: Silva, B., Moro, S. & Marques, C. (2021). Sensing the impact of COVID-19 restrictions from online reviews: The cases of London and Paris unveiled through text mining. In Reis, J. L., Parra López, E., Moutinho, L., and Santos, J. P. M. dos. (Ed.), *Marketing and Smart Technologies. Smart Innovation, Systems and Technologies*. (pp. 223-232). La Laguna: Springer Singapore., which has been published in final form at https://dx.doi.org/10.1007/978-981-16-9268-0_18. This article may be used for non-commercial purposes in accordance with the Publisher's Terms and Conditions for self-archiving.

Use policy

Creative Commons CC BY 4.0

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in the Repository
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Sensing the impact of COVID-19 restrictions from online reviews: the cases of London and Paris unveiled through Text Mining

Bruno Silva¹, Sérgio Moro² and Catarina Marques³

¹ ISCTE – University Institute of Lisbon, Lisbon, Portugal
brmsa@iscte-iul.pt

² ISCTE – University Institute of Lisbon, Lisbon, Portugal
sergio.moro@iscte-iul.pt

³ ISCTE – University Institute of Lisbon, Lisbon, Portugal
catarina.marques@iscte-iul.pt

Abstract. This study aims to understand how the COVID-19 pandemic affected the hotel sector and to identify the current traveler demands. The traveler's reviews were analyzed based on sentiment analysis and a guest satisfaction model was also proposed, demonstrating a data mining approach within tourism and hospitality research. Given its popularity, TripAdvisor was the chosen platform for collection of hotel reviews in London and Paris. Text data were extracted from reviews made in two time periods, before and during the COVID-19 pandemic. The sentiment and specific aspects highlighted by travelers were compared between each period.

Keywords: Text mining, Sentiment Analysis, Tourism, hotel traveler's online reviews, COVID-19 pandemic.

1 Introduction

Over the past years, we have witnessed a notorious increase in the number of text reviews made in several online accommodation platforms following the exponential growth in tourism (Chan et al., 2021). However, during the year of 2020 as the COVID-19 pandemic continued to spread, the online reviews amount has completely changed, namely with the drastic decrease of travel research (Uğur & Akbıyık, 2020). COVID-19 pandemic brought a radical change in people's lives with many economic sectors severely impacted. Tourism was one of the most affected business sectors contrasting with the steady growth seen over the years (Chan et al., 2021). For this reason, it is now extremely important to help tourism, and more specifically hotels, to understand how COVID-19 affected the sentiment of travelers during their stays. The present research aims to identify a gap on subareas within tourism on which data science has not yet been applied. In more particularly this investigation will focus on data mining on online reviews from the TripAdvisor platform, written during

COVID-19 pandemic, to understand what travelers are seeking and what demands they now have. The TripAdvisor reviews were collected from two capital cities, London and Paris, for being very similar in terms of tourism characteristics, focusing on findings answers essentially for one main question: “How is the effect of the pandemic being perceived by visitors to major European capitals severely affected by COVID-19?”

2 Literature Review

2.1 Sentiment Analysis and the impact of COVID-19

Sentiment analysis is a natural language processing tool that is useful for monitoring Web 2.0 applications, as it can reveal public opinion about numerous issues without requiring satisfaction surveys (Valdivia et al., 2017). The interest in sentiment analysis has increased significantly over the last few years due to the large amount of text stored in Web 2.0 applications and the importance of online customer opinions (Valdivia et al., 2017). As a result, more than 1 million research papers contain the term “sentiment analysis” and various start-ups have been created to analyse sentiments in social media companies (Valdivia et al., 2017).

The spread of the COVID-19 global pandemic has generated an exponentially mounting and volume of data that can be harnessed to improve our understanding of big data management research. It is also applied in the need among scholars, practitioners and policymakers for a better and deeper understanding of a range of analytical tools that could be utilized to better anticipate and respond to such unforeseen ‘black swan’ events and risks. (Sheng et al., 2020).

2.2 Conceptual framing and research hypothesis

Based on other studies that identified a list of dimensions drawn from quantitative features known to influence customer satisfaction under the context of TripAdvisor, the list below (Table 1) was considered, as well as a proposal for a guest satisfaction model to illustrate a data mining approach within tourism and hospitality research.

Table 1. List of considered dimensions for this study

Dimension	Reference
<i>Services</i>	<i>(Chan et al., 2021)</i>
<i>Amenities</i>	<i>(Moro et al., 2019)</i>
<i>Health measures</i>	<i>(Chan et al., 2021)</i>
<i>Hotel facilities</i>	<i>(Moro et al., 2019)</i>
<i>Location</i>	<i>(Chan et al., 2021)</i>
<i>Value/price</i>	<i>(Chan et al., 2021)</i>
<i>Cleanliness</i>	<i>(Moro et al., 2019)</i>
<i>Type of travel</i>	<i>(Moro et al., 2019)</i>
<i>Hotel prestige</i>	<i>(Moro et al., 2019)</i>
<i>Seasonality</i>	<i>(Moro et al., 2019)</i>

Restuputri et al. (2021) published a research comparing the staff service quality, operational, technical logistics service providers, with customer satisfaction and loyalty during the COVID-19 pandemic. In this study they refer that an employee must be reliable, punctual and careful at work. An employee must also have effective communication skills, be courteous and ready to serve. The quality of operations service from source to customers must be well-coordinated, on time and with appropriate transportation capacity so no damage occurs to the customer's property. In the end he concluded that a good relationship between staff, operational and technical services is crucial for a good perception of customer satisfaction and loyalty to the services provided by hotels.

Based on the above, the main hypotheses tested in this research was:

H1: Hotel guests appreciate safety concerns by hotels and express about them in online reviews.

Big crisis affects the macro-environment that brings big changes in customer behaviors and hotel performances (Chan et al., 2021). The outbreak of COVID-19 has brought changes in guest experiences at hotels as customers usually describe their experiences and feedbacks after their stays in the form of user-generated content, such as online reviews which are then used by practitioners to understand the nature and structure of guest experiences (Chan et al., 2021). Since hotels are customer-centric that should keep up with customer preferences and requirements, it is essential to understand the impact of COVID-19 and track the changes of guest experiences brought by the same (Chan et al., 2021).

3 Methodology

The aim of this investigation is to propose a guest satisfaction model by seeking any available data and to illustrate a data mining approach within tourism and hospitality research. One of the methods utilized was based on the identification and analysis of a list of specific dimensions drawn from quantitative features known to influence customer satisfaction under the context of TripAdvisor (Moro et al., 2019).

The method used for this investigation is based in a problem resolution paradigm that contains several activities (Alexandre, 2017):

Problem identification and motivation was the first step which defines that the artefact must be viable, that is, it must provide a solution to the research problem. In the presented case, it will be produced in the form of a method to study the impact of COVID-19 restrictions on the TripAdvisor reviews from 2 foreign locations very similar in terms of tourism characteristics. Chosen the platform for the data extraction it was necessary to define the solution objectives, which must be achievable, in addition of the need to understand if this analysis could identify a research gap in the utilization of data mining in tourism.

The third step, Development, involved the current creation of the artefact defined in the first step, that is, the development of a method that allows the analysis and comparison of both hotel reviews. For the sake of the extraction a pre-existent coding script was chosen to perform the scraping, that contained several features to quickly setup and extract data from a website. In this phase it was necessary to understand if the objective and its influencing factors could be translated into data features and instances (Moro et al., 2019). As soon as this was achieved the data needed also to be prepared (data cleaning) for model training.

Next to development was the Demonstration, which was used to verify the effectiveness of the solution to the problem identified. To achieve this goal, graph plots and word clouds or any other most appropriate visualization methods were used in addition to and a pre-build script to plot the most common words from the hotel reviews.

The qualitative measurement of how well the created artefact supports the solution to the problem corresponded to the next step, the Evaluation, which involved the use of the model previously trained to make predictions versus observations (Moro et al., 2019).

And lastly, the communication of the results, whose purpose helped highlight and identify the importance of the problem through the dissemination of the developed artefact, that is, its usefulness and its relevance to other researchers.

4 Results

The discussion of the output results will help clarify of what were the major differences within the customer reviews between the pre and during COVID-19 restrictions on equivalent time periods. In addition, a comparison between capital cities is provided.

Fig. 1 represents the number of reviews extracted per year and city. A significant reduction can be immediately seen from 2019 to 2020 in both cities. In case of London there is, in average, a reduction of 86% on the reviews and Paris with a 95% of decrease. Nevertheless on 2020 London had a slight increase between June and October mainly due to ease of lockdown restrictions in that period (Institute for Government, 2019), followed then by a second decrease as the number of deaths from Covid-19 disease were rapidly rising, as per Fig. 2.

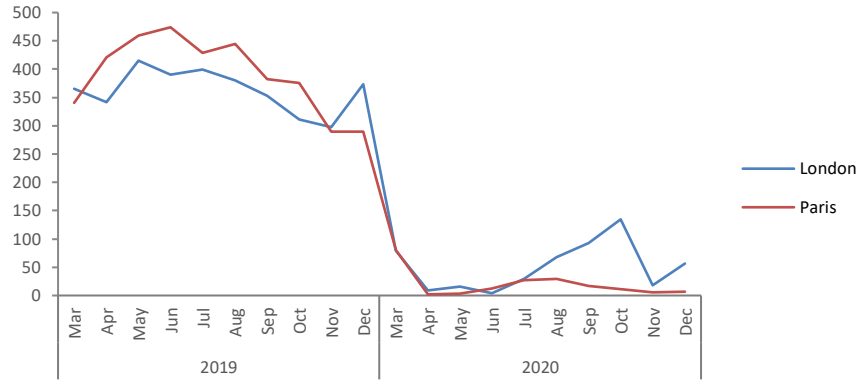


Fig. 1. Distribution of reviews during 2019/2020 periods

Fig. 2 provides a more detailed insight about the evolution of reviews of TripAdvisor and the number of Covid-19 cases and deaths, reported in each city per month. The data from 2019 is also showed to have a fair comparison in the homologous period. In 2020, an increase in the number of reviews, in London, can be seen during the peak of Summer. An increase is also noticeable in Paris but with much less evidence. This difference is related to the severity of the restrictions implemented on each city which were higher in Paris.

In fact, this chart allows to determine the periods where each government applied and eased the pandemic lockdown restrictions. In the case of London, the ease of restrictions started on 23 June 2020 and on 31st of October, England announces second national lockdown (Institute for Government, 2019). During this period the number of reviews done on TripAdvisor rose compared with the first wave of cases.

On a same analogy, in Paris, the government eased the restrictions on 14th of June 2020. On 20th of August France announced that the pandemic was again on the rise (Aurore & Blue, 2020). Due to this, there was also a slight increase in TripAdvisor reviews but with much less expression, compared to London.

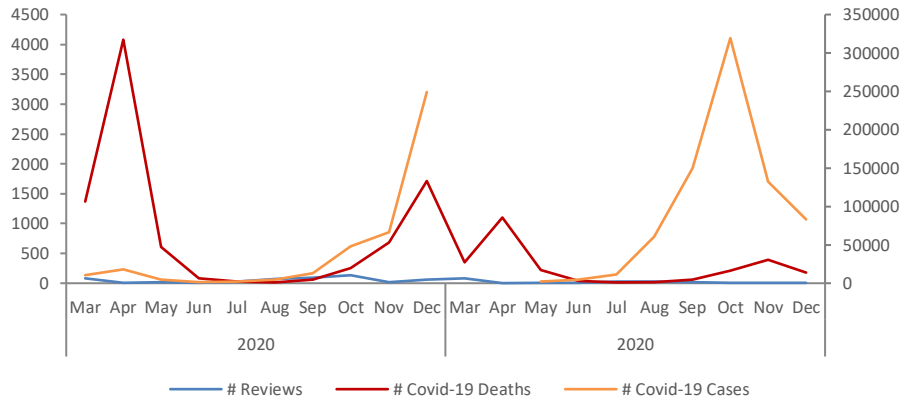


Fig. 2. Distribution of reviews and Covid-19 cases/deaths during the 2020 period

Fig. 3 show the difference in terms of the nationality of the TripAdvisor reviewers. The total number of reviewers categorized by nationality and the percentage based on the total in each year are compared. The scenario of London, in 2020, is an example of expected decrease in the total number of reviewers (87%) and when comparing in terms of nationality a significant increase is noticeable on the national tourists of about 40%.

Paris, on the other hand, demonstrates that the city is more visited by the international tourists than those living within borders, in 2020. This behaviour could be explained by the fact of the collected data being filtered in the English language only. Also, this last point can explain the significant decrease in the written reviews on TripAdvisor (95%) as it is very likely that, for Paris, there are even more reviews written in French.

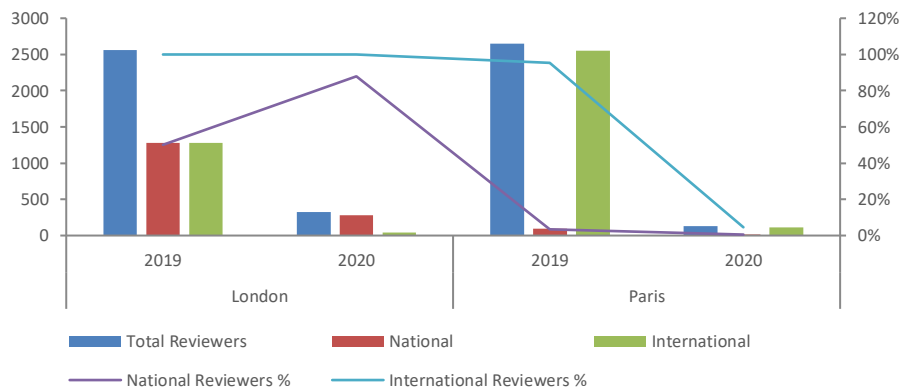


Fig. 3. Comparison between National and International reviewers per city and year

The sentiment analysis part was also included in the mining process of TripAdvisor's extracted data. Fig. 4, Fig. 5 and Fig. 6 present three different perspectives on the analysis of guest satisfaction based on the dimensions chosen for the study that were already discussed in previous chapters.

To evaluate the guest satisfaction, in each review, the sentiment words were drawn, and their scores were calculated. A score higher than zero was considered a positive review, a score equal to zero a neutral score and finally a score less than zero a negative one. For example, reviews with the words "helpful", "greeting", "friendly", "amazing" got the best scores and reviews with word combinations like "excellent space optimization" and "real pleasant surprise" also were classified with best positive scores. After the identification of the key words for each review, a new classification was done using the list of ten dimensions shared in the study. An average was then calculated per each dimension based on positive and negative perspectives.

As per Fig. 4 and Fig. 5, almost every dimension was classified with a positive sentiment, in both years. It is possible to see that in Paris during 2020, seasonality negatively impacted the reviewers when scoring their stays. This may be justified by the pandemic situation lived during this time. Health measures is another dimension affected by this reason getting 42% of negative reviews. In the case of London, curiously, was not possible to retrieve any feedback referring both seasonality and health measures dimensions.

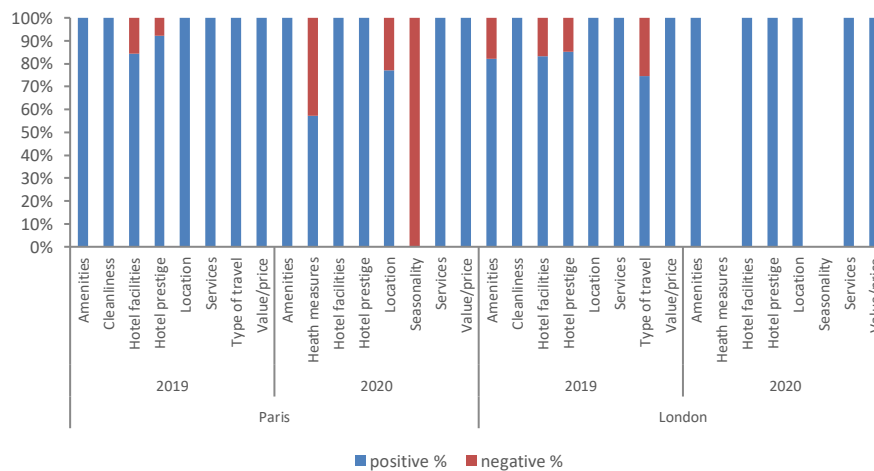


Fig. 4. Sentiment perception distribution by dimension over total number of reviews

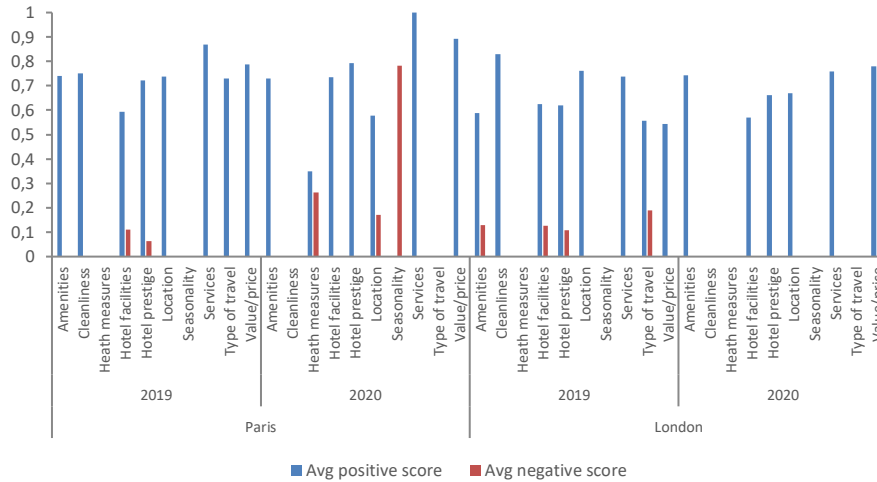


Fig. 5. Sentiment perception average score by dimension, city and year over total number of reviews

Fig. 6 shows another perspective of the dimension analysis. As already referred seasonality clearly stands out as the most negative perceived dimension, in 2020. Health measures were also classified with low review scores. In 2019 the dimensions with poorer review sentiment were related to amenities, hotel facilities and prestige and the type of travel of the guests. In the other hand, dimensions like the services and value/price were the ones with better sentiment classification, in both cases of Paris and London.

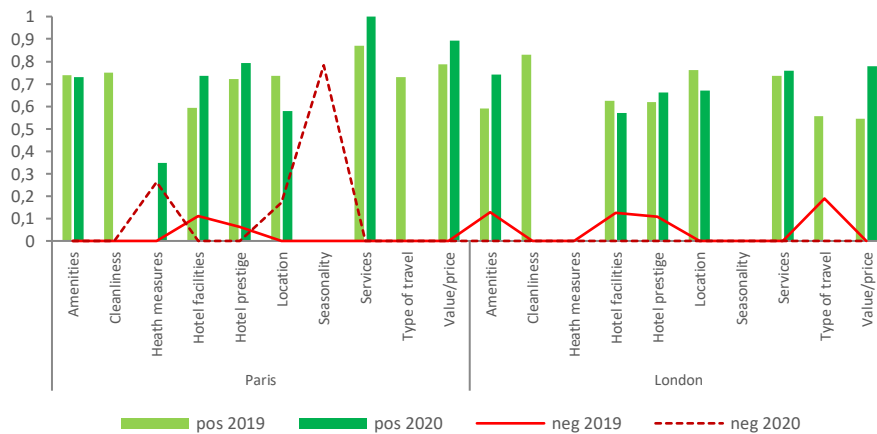


Fig. 6. Sentiment perception by dimension and year over total number of reviews

Table 2 contains a summary of the analysis of each of the dimensions discussed.

Table 2. Effects on customer satisfaction

Effect on customer satisfaction		
Dimension	Paris	London
<i>Services</i>	<i>In general, services have a positive effect on the guests; 2019 – the best rated dimension; 2020 – the worst rated.</i>	<i>The dimension has also a positive perceived value on customer; 2019 – best rated than in 2020 (with a neutral feedback).</i>
<i>Amenities</i>	<i>74% of positive feedback in both years. No negative effect perceived by guests.</i>	<i>2020 – the second-best rated dimension; 2019 – the second worst rated dimension. Nevertheless, the average positive rating is greater than the negative feedback.</i>
<i>Health measures</i>	<i>2019 – top negative rated dimension; 2020 – second negative rated dimension along with Cleanliness.</i>	<i>No feedback extracted.</i>
<i>Hotel facilities</i>	<i>Third negative dimension in 2019; Mostly rated positive in 2020.</i>	<i>Positive rating in general.</i>
<i>Location</i>	<i>Increase of 17% of negative rating value in 2020 compared to 2019.</i>	<i>100% positive feedback.</i>
<i>Value/price</i>	<i>Top positive rated dimension in 2020 and in overall.</i>	<i>Second best positive rated dimension.</i>
<i>Cleanliness</i>	<i>Second most negative rated dimension</i>	<i>Top positive rated dimension in 2019.</i>
<i>Type of travel</i>	<i>More decisive for travelling in 2019.</i>	<i>More decisive for travelling in 2019.</i>
<i>Hotel prestige</i>	<i>Mostly rated positively in 2020.</i>	<i>Positive rating improvement from 2019 to 2020.</i>
<i>Seasonality</i>	<i>Mostly negative perceived by guests in 2020.</i>	<i>No feedback retrieved from analyzed guests reviews.</i>

5 Conclusion

The main objective of this study was to understand what travellers are seeking nowadays during the current pandemic that we are all living. From the results and analysis done on the data mining process it was possible to understand some aspects of the behaviour, profile and most importantly how guests are now perceiving the effect of COVID-19 pandemic in their hotel stays in two major capital cities being this the major research question of this investigation.

A guest satisfaction method was proposed demonstrating that most of the reviews (55%) done on TripAdvisor on 2020 showed a positive perception, despite the COVID-19 pandemic when comparing to 2019.

Some limitations were also felt during the analysis and data extraction that could not allow for a deeper investigation on some interesting hypothesis. The need of travel insurance has raised with the pandemic or the existence of a full refund option in case of travel cancelling could be some examples. To achieve this, it could be worth to perform a deeper analysis into the TripAdvisor reviews data such as extracting the day guests book their travel. Despite not being a mandatory field, it could lead to the testing of other hypotheses such as if the fear of confinement leads visitors to book their travels only on the same day they will actually travel.

References

- Alexandre, I. M. (2017). *Design Science Research*. 1–19.
- Chan, I. C. C., Ma, J., Ye, H., & Law, R. (2021). A Comparison of Hotel Guest Experience Before and During Pandemic: Evidence from Online Reviews. In *Information and Communication Technologies in Tourism 2021* (pp. 549–556). Springer International Publishing. https://doi.org/10.1007/978-3-030-65785-7_52
- Moro, S., Esmerado, J., Ramos, P., & Alturas, B. (2019). Evaluating a guest satisfaction model through data mining. *International Journal of Contemporary Hospitality Management*, 32(4), 1523–1538. <https://doi.org/10.1108/IJCHM-03-2019-0280>
- Restuputri, D. P., Indriani, T. R., & Masudin, I. (2021). The effect of logistic service quality on customer satisfaction and loyalty using kansei engineering during the COVID-19 pandemic. *Http://Www.Editorialmanager.Com/Cogentbusiness*, 8(1), 1906492. <https://doi.org/10.1080/23311975.2021.1906492>
- Sheng, J., Amankwah-Amoah, J., Khan, Z., & Wang, X. (2020). COVID-19 Pandemic in the New Era of Big Data Analytics: Methodological Innovations and Future Research Directions. *British Journal of Management*, 1467-8551.12441. <https://doi.org/10.1111/1467-8551.12441>
- Uğur, N. G., & Akbıyık, A. (2020). Impacts of COVID-19 on global tourism industry: A cross-regional comparison. *Tourism Management Perspectives*, 36, 100744. <https://doi.org/10.1016/j.tmp.2020.100744>
- Valdivia, A., Luzón, M. V., & Herrera, F. (2017). Sentiment Analysis in TripAdvisor. *IEEE Intelligent Systems*, 32(4), 72–77. <https://doi.org/10.1109/MIS.2017.3121555>