The impact of COVID-19 on the Brazilian food service industry: topic modelling of online reviews

Abstract:

The purpose of this paper is twofold. First, it is to discover the underling topics that characterize the customer relationship with restaurants in Brazil. Second, it is to explore the changes in this relationship caused by the necessary actions to mitigate the COVID-19 pandemic. To reach this purpose, 57,381 online reviews of 1,066 restaurants in four Brazilian cities were collected from TripAdvisor. To analyze this data, this research employed text mining and natural language processing techniques, such as topic modelling with Latent Dirichlet Allocation (LDA) and sentiment analysis. Results present a list of the 15 most relevant topics that characterize the customer relationship with restaurants in Brazil. These topics were than divided into five dimensions: Food type, Service, Food quality, Cost benefit, and Place. Due to the COVID-19 pandemic, the number of reviews reduced by 93.27%, comparing with the same period of 2019. In the pandemic period, the four most relevant topics are 'delivery', 'employees', 'experience', and 'waiter service'. This research is among the first to make use of text mining and topic modelling altogether with sentiment analysis to examine consumers perceptions of Brazilian food service industry, especially to understand the impacts of the COVID-19 pandemic on this sector. The results clearly present the most relevant areas that consumers express interest in restaurants, showing paths for managers facilitate and improve their restaurants operations.

Keywords: topic modelling, food service, sentiment analysis.

1. Introduction

The out-of-home food sector is one of the pillars of Brazilian economy. With an annual expansion of around 10% in 2019, bars and restaurants generate about 450 thousand new job opportunities per year. Most places are for waiters - 250,000 new jobs annually (SEBRAE, 2019). Moreover, in 2019, the food service industry represented 33% of total food retail sales in Brazil (Statista, 2019). The international pandemic of the new coronavirus has led to an interruption of society' daily activities, mainly due to the obligation of social distancing (WHO, 2020a, 2020b). Among the measures undertaken to stop the disease spread in Brazil, was the shutdown of bar and restaurants. The fact that these companies had to adapt themselves to keep working by selling its products through delivery systems, has changed the way consumers buy these foods. Instead of going to the restaurant looking for an experience beyond the food, orders are now made to be taken away.

Thus, the main purpose of this paper is to explore the changes in the relationship between restaurants and their customers, caused by the necessary actions to mitigate the COVID-19 pandemic. To achieve this objective, this study collected and analyzed 57,381 textual reviews of 1,066 restaurants in four Brazilian cities. Online reviews are full of customer compartmental insights and can bring unique opportunities to companies as well as academic researchers. Online reviews can be understood as a natural, perceptive, and passionate feedback provided by consumers to other consumers and to the firm. These reviews are universally accessible, usually for free. The hefty amount of data contained in the nonstop stream of online reviews over time can offer practical

input to enhance traditional research for detecting important problems (Guo, Barnes, & Jia, 2017; Luo, Vu, Li, & Law, 2020).

Incorporating new data sources and analytic tools can help to reduce this complexity, and can bring insights to customers experience and customer journey research and practice (Lemon & Verhoef, 2016; Wedel & Kannan, 2016). The world is producing more data than ever before, and it is estimated that by 2020, it will be produced 44 zettabytes worth of data every day. For marketers charged with delivering better customer experiences this means of a plethora data to work through to extract actionable insights (Morgan, 2017). Using data collected on customers and prospects customers outside the firm environment can offer better descriptions and predictions of consumer choice (Timoshenko & Hauser, 2019; Wedel & Kannan, 2016). In this sense, online reviews act as an important source of knowledge for firms to achieve a better comprehension not only of their customers but also of the conditions of the market they operate. Moreover, online reviews provide an abundance of data in both structured form (e.g., scaled ratings) and unstructured form (e.g., open-ended responses, reviews) (Balducci & Marinova, 2018; Berger et al., 2020; Kannan & Li, 2017; Korfiatis, Stamolampros, Kourouthanassis, & Sagiadinos, 2019). To analyze this massive amount of data, analytic tools are paramount. Recent developments in analytics, such as machine learning, topic modelling, and natural language processing are particularly useful in this analysis (Nagji, Venkatesan, Trzcinski, & Craddock, 2018), with text mining and other emerging technologies are potentially better ways to measure and manage the reviews (Lemon & Verhoef, 2016; Wedel & Kannan, 2016).

The hospitality industry is specially affected by online reviews due to its information intensive nature, and due to the proliferation and dissemination of specialized websites, such as Yelp! and TripAdvisor (Chung & Koo, 2015). Hospitality refers to a variety of businesses linked to leisure. It includes the accommodations sector (e.g., hotels, hostels, resorts), the food and drink sector (e.g., restaurants, coffee shops, nightclubs) and the travel and tourism sector (e.g., attractions, tour operators, car rentals) (Brotherton, 1999; Hemmington, 2007). Most of the previous research that studied the hospitality industry through online reviews focused on the accommodation sub-sector.

For instance, the research of Liu et al. (2017), gathered 412,784 user-generated ratings of 10,149 hotels from five Chinese cities on TripAdvisor, and found that foreign tourists differ substantially from local tourists in terms of their emphasis on the roles of various service quality attributes in forming their overall satisfaction rating for hotels. Guo et al. (2017) identified the key dimensions of hotel service quality using a data mining approach. Their data set includes 266,544 online reviews for 25,670 hotels located in 16 countries. In their research, Ju, Back, Choi, and Lee (2019) adopted a mixed method approach to capture the various service aspects of Airbnb by collecting and analyzing 16,430 online reviews. Their findings suggest that Airbnb has multiple service quality attributes associated with website, host, and facility that produce distinctive effects on consumer satisfaction.

This research helps to filling this disparity, by studying the restaurant industry in Brazil. Regarding the COVID-19 pandemic, Brazil is also a relevant subject as it is the second most affected country (in absolute numbers) in both number of infection and number of deaths (Statista, 2020). This paper can also provide novel insights to enhance the effectiveness of designing the business operations. Insights from the customer's perspective is helpful in exploring what can go wrong and how to put it right (Chheda, Duncan, & Roggenhofer, 2017). For the decision makers and managers, this study provides the main topics in which consumers are concerned in their restaurant consumption. The analysis will also provide the level of importance of each topic, so managers can allocate their resources more efficiently. By focusing on written data from each

customer online review, specialists can extract precious insights about the touchpoints from the customer's perspective, recognizing which touchpoints are working well and those demanding attention. Moreover, firms can segment customers based on their reviews. Firms can use these insights to detect the underlying factors triggering pain for customers and then prioritize which root causes need attention. Finally, managers can use the results do develop more effective customer journeys and improve their operations.

2. Research methodology

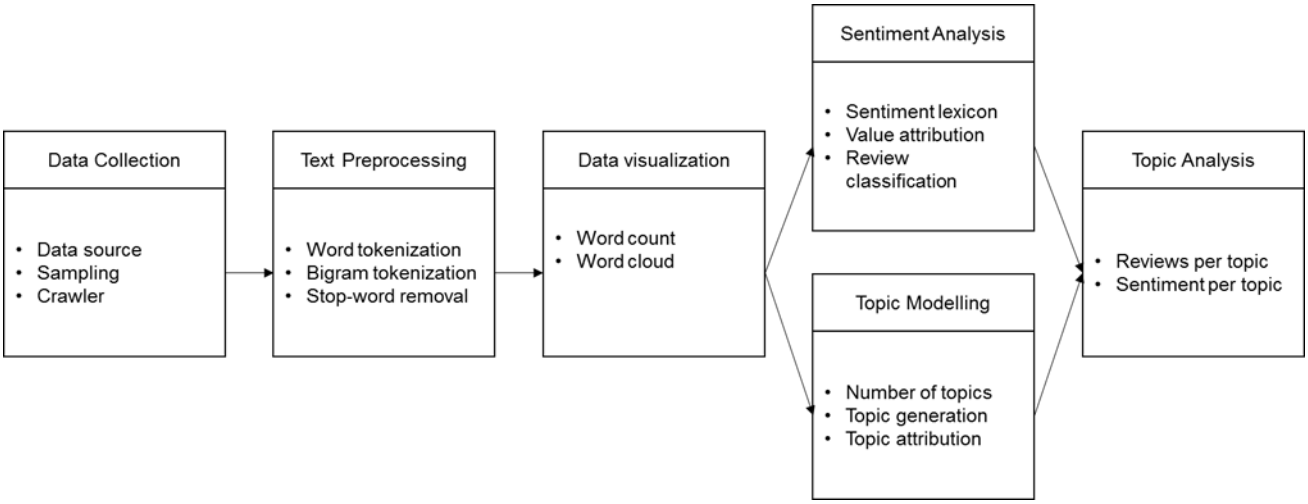
This research employs several techniques to collect, analyze and explore the online reviews of Brazilian restaurants, and aims to answer the following questions:

(1) What are the key topics that Brazilian food service customers express trough online reviews?

(2) How these topics changed due the impacts of the COVID-19 pandemic?

The framework presented in Figure 1 summarize the steps followed in this research: data collection, text preprocessing, data visualization, sentiment analysis, topic modelling, and topic analysis, as suggested by Berger et al. (2020). All analyses were conducted using the software R (R Core Team, 2020). Details on each step are presented the next topics.

Figure 1
Methodological framework



2.1. Data collection

Restaurant reviews are available on a variety of online platforms, such as Facebook, Google Reviews, 'Yelp!', and TripAdvisor. For this study, TripAdvisor was chosen as data source, as its reviews were proven reliable and have been extensively exploited in previous research (e.g., Luo et al., 2020).

The WebCrawler R package 'Rcrawler' (Khalil, 2018) was used to automatically scrap the reviews and information of restaurants of four Brazilian capitals: Belo Horizonte, Rio de Janeiro, São Paulo, and Vitória. Collected information included: textual review; review title; reviewer

profile; restaurant name; and restaurant location. Data collection was conducted in July 2020, and the final sample is composed of 57,381 textual reviews of 1,066 restaurants in four Brazilian cities, as presented in Table 1. The sample is composed of daily data, spanning from January 1, 2019 to June 30, 2020. Figure 2 presents the number of reviews per day. All collected reviews were written in Portuguese, and in order to better present the results, the words were manually translated to English.

Table 1:
Sample Profile

City	Number of reviews		Number of restaurants		Review per restaurant
Belo Horizonte	12,340	22%	294	28%	41.97
Rio de Janeiro	15,718	27%	285	27%	55.15
São Paulo	26,961	47%	280	26%	96.29
Vitória	2,362	4%	209	20%	11.30
Total	57,381		1,068		53.73

2.2. Data preprocessing

Natural language processing of textual data such as online reviews require some preprocessing in order to tidy and transform texts into machine-friendly formats (Berger et al., 2020; Blei, Ng, & Jordan, 2003; Liu et al., 2017). First, the reviews were tokenized into single words. All punctuation, numbers, and stop-words (commonly used words in a language, such as "an", "and" and "with") were removed, as they appear repeatedly and do not give considerable insight to the analysis. The list of stop-words for Portuguese, as well as the preprocessing were conducted with the R package 'tm' (Feinerer & Hornik, 2019). This process converted the unstructured texts into structured data, ready to be worked with. For example, an initial review appeared as:

> "Very good meat! Despite liking the meat, this time I ate the cod with rustic potatoes and broccoli and loved it! Worth checking out! And the wine menu is also good!"

After pre-processing the review, it became:

> "good meat despite liking meat time ate cod rustic potatoes broccoli loved worth checking wine menu good"

2.3. Data analysis

Two main analyses were conducted in this paper: sentiment analysis, and topic modelling trough Latent Dirichlet Allocation (LDA, Blei, Ng, & Jordan, 2003). Sentiment analysis is a text mining methodology that is used to measure customer emotive polarization (positive and negative feelings) attitude in texts (Feldman, 2013; Taboada, Brooke, Tofiloski, Voll, & Stede, 2011). In the case of this paper, a lexicon of positives and negative words in Portuguese, provided by the R package 'lexiconPT' (Gonzaga, 2017), was used to generate a score for each review. The lexicon attributes a positive (+1), negative (-1), or neutral (0) value for each word in a review. The sum of these values provides an overall sentiment score for a review, thus the higher the value, more extreme are the sentiment expressed in the review.

For example, a review with score +4 is:

> "Very good! A very cool bar, with live music, pleasant atmosphere, good food and good service. We ordered the great portion of filet baits with 3 sauces (R$ 47.00) and also tasted the excellent Kud God Save the Queen beer (R$ 12.00)."

And with score -4 is:

> "Disappointed. It is sad to go back to a restaurant and realize that it turned bad. The atmosphere is still beautiful, but the pizza is cold and mediocre, and the service is inattentive and negligent."

A topic model, in turn, is a type of probabilistic model for uncovering the abstract topics that arise in a collection of documents (Blei et al., 2003; Guo et al., 2017). The most common method for topic modelling is LDA, that is, a machine learning tool that enables the discovery of underlying topics from huge volumes of unstructured text data. In LDA, each unique word in a text have a probability value of belonging in a topic. In this sense, an LDA topic is theoretically formed by all unique words, with different probability values. Similarly, a text is also formed by different topics with different probability values, and each review will have its own combination of percentage of topics. The R packages 'tidytext' (Julia Silge & Robinson, 2017), 'tm' (Feinerer & Hornik, 2019), and 'topicmodels' (Grün & Hornik, 2011) were used for the LDA analysis.

The results of the analysis are presented in the next topic. Fists, a summary of the data with its visualizations is shown. Next the results of the sentiment analysis and of the topic modelling are presented. Finally, a joint analysis of these two analyses is conducted.

3. Results

3.1 Visualization

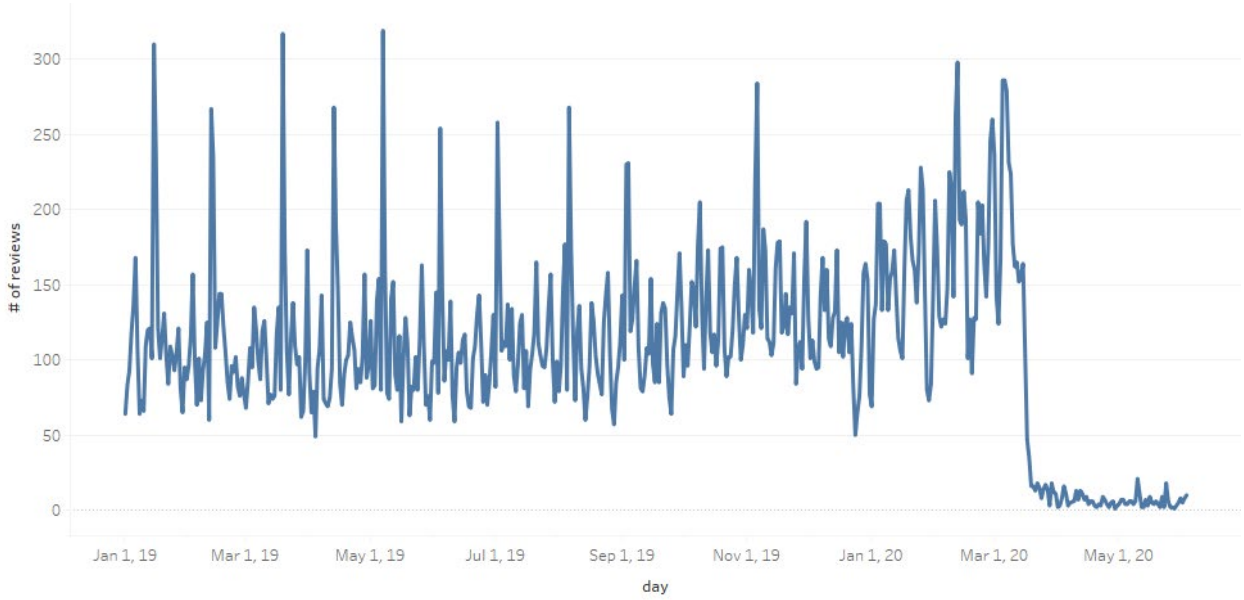
A first result is the plot of the number of reviews per day in the sample, which is presented in Figure 2. The peaks in the graphic matches with national and local holidays, as well as weekends. The last peak in the graphic is on March 06, 2020. From this day, a major drop in the number of reviews starts, stabilizing in the 'new normal' from March 19, 2020. The average of review per day for the hole sample is 106.33 reviews. The average of 2019 is 114.57, while in 2020 is 86.80.

Comparing the period of the pandemic impact in the number of review (from March 19, 2020 to June 30, 2020) with the same time period of 2019, it is possible to better understand this impact. For this period in 2019 the average was 107.54, when for 2020 is 7.23 reviews per day. The total number of reviews for January and February 2020 were significantly higher than 2019, and even March 2020, the month that the World Health Organization declared that the world as living a pandemic, was closer to the number for March 2019. This result indicates that if the pandemic had not happened, the number of reviews in 2020 could be larger than 2019.

Entering the text analysis, the most common word counting visualization is 'word cloud'. In this plot, the bigger the size of a word, more times it appears in the sample. The R package 'wordcloud' (Fellows, 2018) was used to generate the 'word cloud'. Figure 3 shows the word cloud for the pre-pandemic period and for the pandemic period. While some words remain the same, there is a clear difference among the clouds. For the total sample, words such as 'experience', 'dishes', 'nice', and 'best' are the most frequent, while for the pandemic cloud novel words as 'delivery, 'house', 'order', and 'iFood' (a food delivery platform) emerged. This result shows that, during the pandemic, customers have been less concerned with the experience provided by the restaurant and have focused on the delivery aspect. It is important to notice that a possible scenario

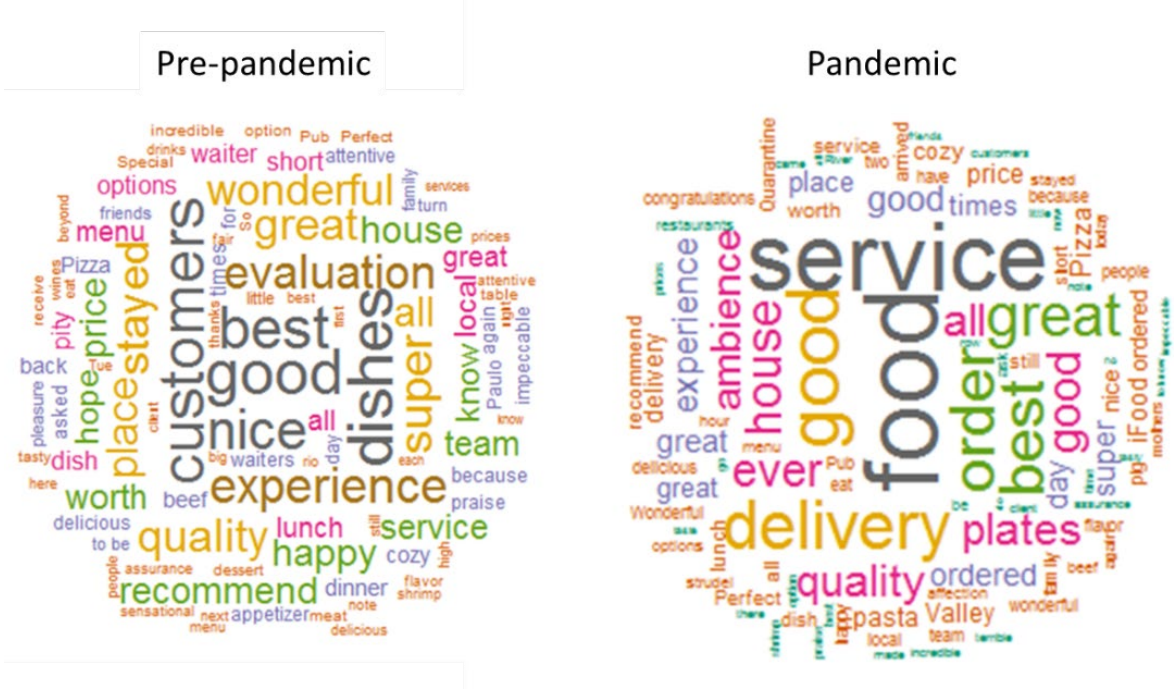
for most of these restaurants, is that they did not delivered food before the pandemic and had to adapt their operations to fit this new demand and to keep on selling.

Figure 2
Reviews per day



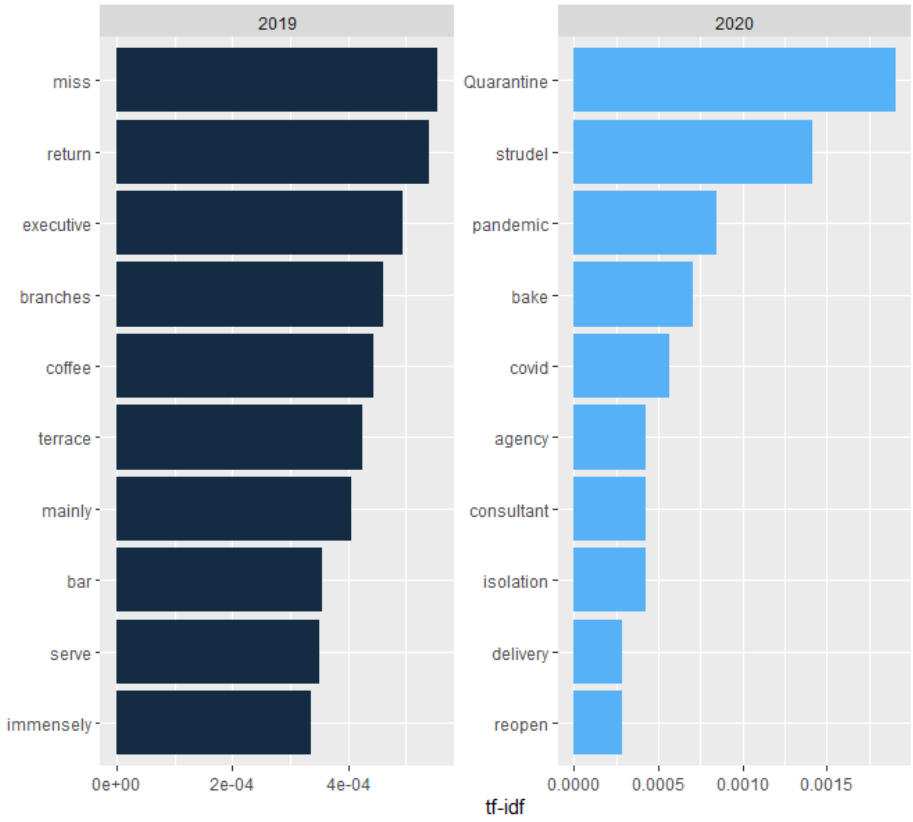
The trend of distinct count of Review trabalhada for Date Day.

Figure 3
Word cloud



The next analysis conducted was the term frequency–inverse document frequency (tf-idf). The idea of tf-idf is to discover relevant words of each group by diminishing the weight for commonly used words and increasing the weight for words that are not extremely used in a collection. The calculation of tf-idf tries to find the words that are common in a text, but not excessively common (J Silge & Robinson, 2016). To conduct this analysis, the data was divided into two groups: April, May. and June, of 2019 and 2020. The results presented in Figure 4 confirm the previous analysis and show the difference of exclusive words between the two periods. Unique words for the pandemic period are 'quarantine', 'pandemic', '*covid'*, 'isolation', 'delivery' and 'reopen'. It is clear that all these words are related to the COVID-19 pandemic.

Figure 4
Tf-idf unique words



3.2 Sentiment analysis

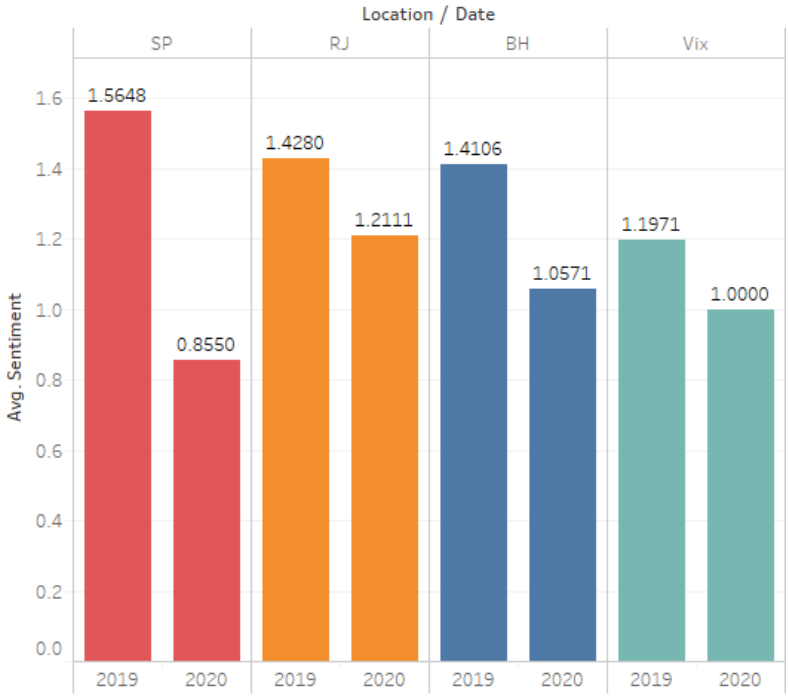
The sentiment analysis conducted provide interesting insights. First, the average value of sentiment of the total sample is 1,506. This means that the overall sentiment of the consumers toward the restaurants is positive. The outline of the sentiment analysis is presented in Table 2. For the pandemic period, the average is lower than the previous periods.

Next, Figure 5 reveals the average of the sentiment analysis for the pandemic period (April to June) of 2019 and 2020, separated by city. It is clear that the sentiment toward the restaurants reduced in the pandemic period, especially for São Paulo. This can be a reflex of the lack of experience of the restaurants with delivery systems, or that the restaurants are no capable to deliver the same value to the consumer.

Table 2
Sentiment analysis

Sentiment	Total sample	2019	2020	2019 (Apr-Jun)	2020 (Apr-Jun)
Min	-20	-10	-20	-5	-4
1st Q	0	0	0	0	0
Median	1	1	1	1	1
Mean	1.506	1.506	1.508	1.477	0.984
3stQ	2	2	2	2	2
Max	60	60	45	40	7

Figure 5
Sentiment per city (pandemic period)



Average of Sentiment for each Date Year broken down by Location. Color shows details about Location. The marks are labeled by average of Sentiment. The data is filtered on Date Month, which keeps April, May and June.

3.3 Topic modelling

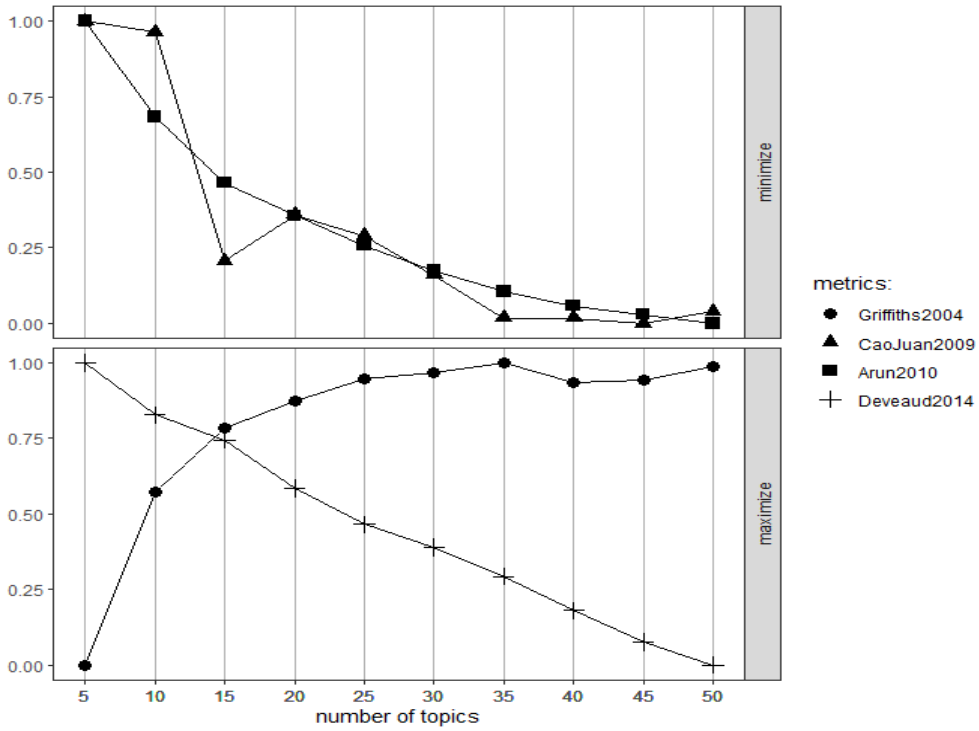
To answer the research questions, two topic modelling procedures were conducted, for the pre-pandemic period, and for the pandemic period.

3.3.1 LDA for the pre-pandemic period

Topic modelling with Latent Dirichlet Allocation require the specification of the number N of topics to be calculated. The number of topics can influence the overall quality of the machine learning algorithm results, and so, an appropriate number N should be selected. In this paper a combination of four metrics is used to define the optimum number of topics (Nikita, 2020), as presented in Figure 6. The objective of this analysis is to find the best trade-off between the four metrics, minimizing the two measures in the top-panel (Figure 6), while maximizing the two

measures in the bottom panel. Based on the graphic analysis, N = 15 topics seems to be a good trade-off between the four measures.

Figure 6
Number of topics



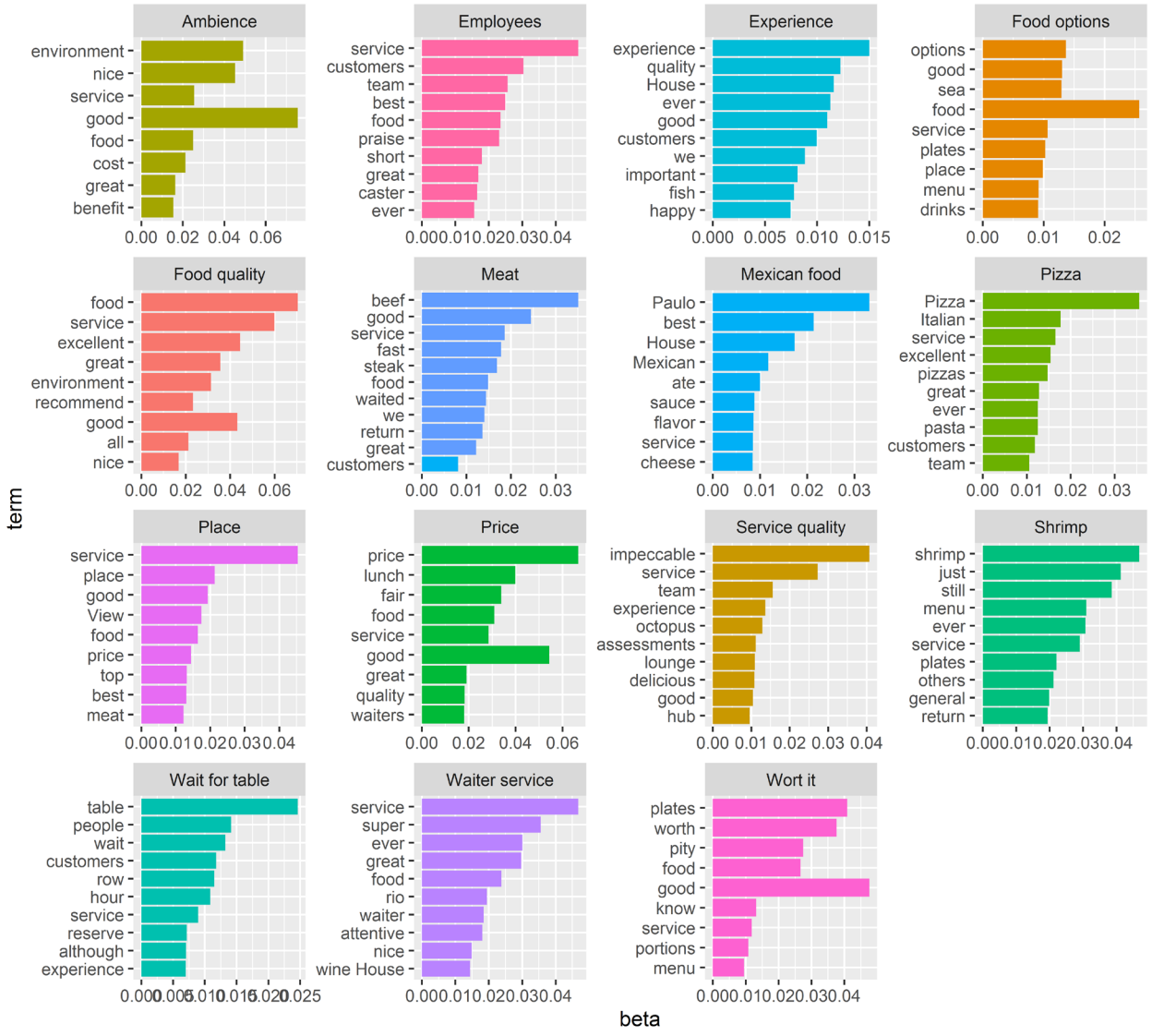
After defining the optimum number of topics, the LDA analysis was conducted. First, about 10% of the total number of reviews were randomly chose and were used as a training set for the LDA algorithm. Figure 7 presents the 15 topics and the ten words with higher probability values for each topic. The topics were labeled based on the analysis of these words. Next, the trained LDA algorithm was used to attribute the topic probabilities for each review of the dataset. The most common topic are meat, place, Mexican food, and waiter service. Figure 8 provides the number of reviews for the 15 identified topics. These 15 topics were than divided into four dimensions: 'food type', 'service', 'food quality', 'cost benefit', and 'place' (Table 3).

3.3.2 LDA for the pandemic period

To test the pandemic effect on online review content, a second LDA algorithm was calculated. Due to the relatively small number of reviews for the pandemic period (n = 382), the number of topics was calculated again. This time, four topics was considered a good trade-off between the metrics. Thus, N = 4 was used to calculate de LDA.

The LDA topics resulted are presented in Figure 9. As expected, delivery now emerged as an important topic, along with food quality, place, and service. Three of the four topics generated are similar to topics of the LDA analysis for the whole period. Thus, the main effect that the COVID-19 pandemic exercised in the topics was first the reduction of topic numbers. Topics such as 'employees', 'experience', 'waiter service', and wait for table are no longer in the main topics.

Figure 7
Topics from LDA



3.4 Topic analysis

Finally, each of the individual reviews was associated with a single topic, and it was calculated the sentiment per topic. For all topics, the associated sentiment was positive, and the topics with greatest associated sentiment were Wait for table, Shrimp, Ambience, and Experience. In contrast the topics with lowest sentiment were Food quality, Waiter service, Mexican food, and Beef. Further, the same analysis was conducted, but now the data was segmented by city. The analysis is presented in Figure 10 as well. The city of Vitória has the lowest score for almost all topics, except for Shrimp. São Paulo, on the other hand, has the highest scores for ten of the 15 topics. The Pizza topic presents the greatest difference among the four cities, and it seems that São Paulo has the better pizza. This topic analysis is especially relevant for restaurant managers, as it gives an

overall analysis of the restaurant marketplace and indicates in which areas they need to focus to improve the consumer sentiment toward their establishments.

Figure 8
Number of reviews per topic

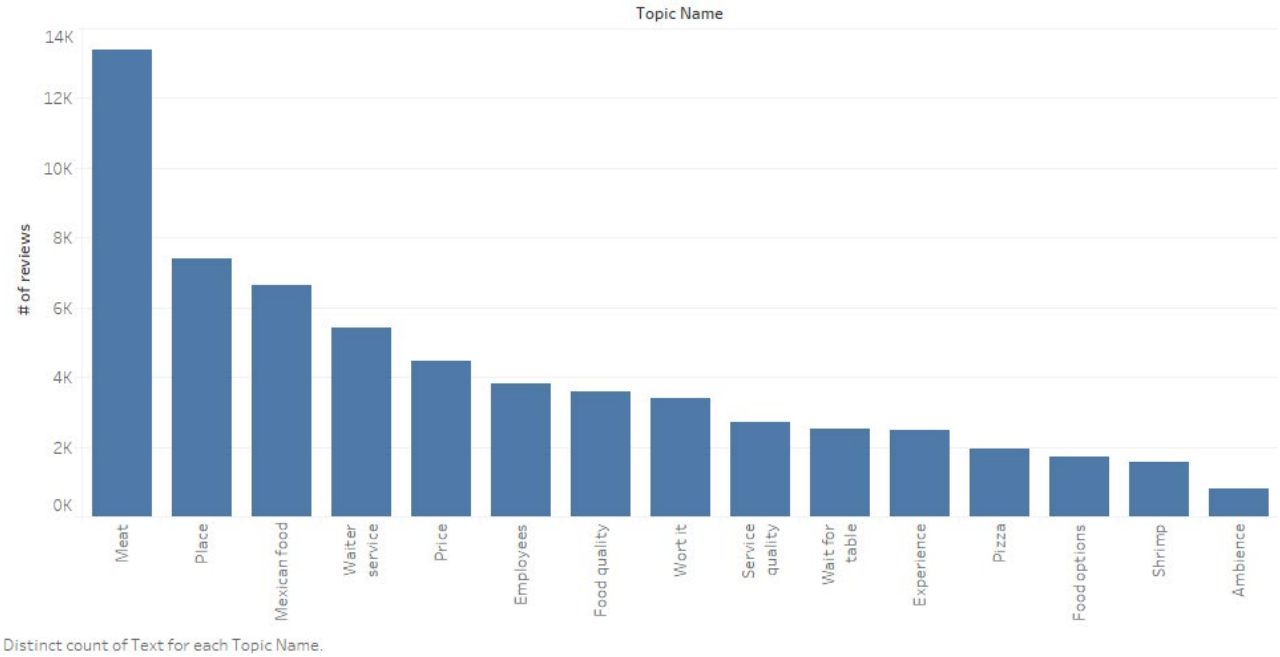


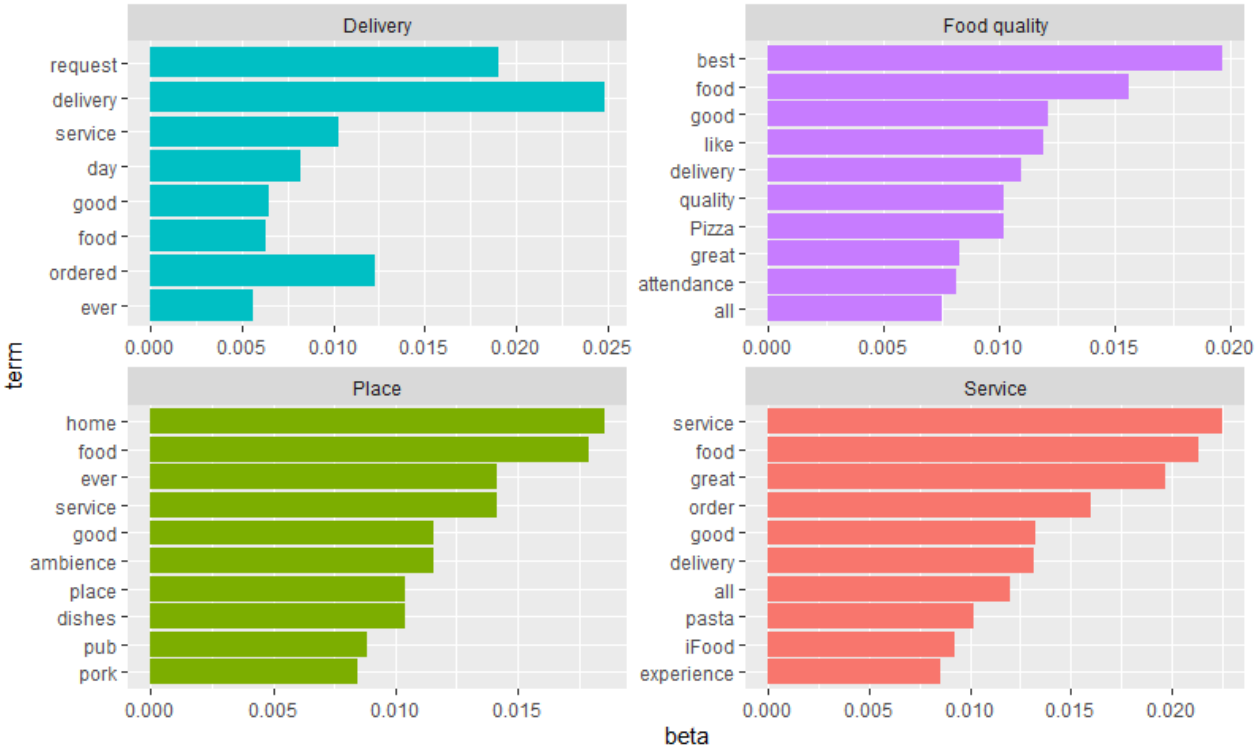
Table 3
Dimensions of topics

Dimension	Food type	Place	Service	Cost benefit	Food quality
Topic	Pizza	Ambience	Employees	Price	Food options
	Mexican food	Experience	Service quality	Worth it	Food quality
	Beef	Wait for table	Waiter service		
	Shrimp	Place			

4. Discussion and implications

This study examined online reviews on Brazilian restaurants, discovering and analyzing the underlying topics related to the consumers experiences in these restaurants, trough topic modelling with Latent Dirichlet Allocation (Blei, Ng, & Jordan, 2003). The results indicated that there are 15 most significant topics related to the consumer experiences in these restaurants, namely Pizza, Ambience, Employees, Price, Food options, Mexican food, Experience, Service quality, Worth it, Food quality, Beef, Wait for table, Waiter service, Shrimp, and Place. These topics were then grouped into five dimensions, namely Food type, Place, Service, Cost benefit, and Food quality. The most frequently mentioned topic is Beef, expressing the preference of the studied consumers for this kind of food. It also can be a reflex of the number of restaurants that sells beef dishes. Other types of food that are popular among these consumers/restaurants, that emerged from the topic analysis, are Mexican food, Pizza, and Shrimp. These topics and dimensions represent the attributes that consumers find most relevant in their relationship with the restaurant (Luo et al., 2020).

Figure 9
Topics for the pandemic period



To complement the topic modelling, a sentiment analysis was conducted. The overall consumer's sentiment in the direction of the restaurants analyzed is positive, as well as the overall sentiment for each of the discovered topics (Berger et al., 2020). The topics with greatest associated sentiment were Wait for table, Shrimp, Ambience, and Experience. The restaurants in the sample appears to excel in these attributes. The topics with lowest sentiment were Food quality, Waiter service, Mexican food, and Beef. In this sense, although the sentiment toward these topics are positive, there are room for improvement in the quality of the food in these restaurants. Additionally, the sentiment per topic were compared throughout the four cities studied. The city of São Paulo has the highest general scores for almost all topics, and in contrast the city of Vitória has the lowest score for almost all topics, except for one. These results show that cultural, economic, and social differences can influence both the performance of the restaurants and the consumer perception on each topic.

Moreover, to investigate the impact of the COVID-19 pandemic on these topics, the topic analysis was also conducted for this period. The results indicated that there are four main topics for this period. Delivery appeared as an important topic, along with Food quality, Place, and Service. Beyond the topics, the COVID-19 pandemic severely affected the Brazilian restaurant business, as shown by the radical decrease in the number of reviews. The number of reviews from April to June in 2019 averaged 107.54 reviews per day, while for the same period of 2020 the average is 7.23, a reduction of 93.3%. Even though the restaurants remained functioning through delivery systems, it seems that it is not enough to keep its consumers. The sentiment analysis revealed that the consumers sentiment toward the restaurants, also reduced in the pandemic period. In this sense, consumers seem to be less satisfied with the restaurant's services than before the

pandemic. This is another sing that the restaurants had difficulties in adapting their operations for the pandemic.

Figure 10
Sentiment per topic per city



This research can provide valuable information for managers of the food service industry. The LDA results altogether with the sentiment analysis clearly presents the most relevant areas that consumers express interest in restaurants, showing paths for managers facilitate and improve their restaurants operations. The findings also of this research suggest that the waiter service and staff are important in influencing the consumer experiences and were associated with lower scores of customer sentiment. Therefore, restaurants should improve the quality of its employees, either by training current employees or hiring trained ones. Extra effort should be employed to improve restaurants' food quality, the topic with lowest associated customer sentiment. Managers could improve the quality of their raw food input and invest in training for its cooks and kitchen staff. Nonetheless, the COVID-19 context changed the socioeconomic environment in which restaurants operate. In this new milieu, whereas the consumer orders their food through delivery or take-away, consumers continue concerned with the food quality, service, and place of the restaurant, and thus managers can improve these aspects in order to mitigate their losses. However, a novel topic emerged in this pandemic context. Consumes are valuating the delivery aspect, and managers should keep up with this new demand.

This research is among the first to make use of text mining and topic modelling altogether with sentiment analysis to examine consumers perceptions of Brazilian food service industry, especially to understand the impacts of the COVID-19 pandemic on this sector. The findings strengthen previous research on food service industry, in particular the studies of consumer behavior and service management and can bring insights on how consumers are dealing with the pandemic. Despite the contributions of this paper, an important question remains unanswered. When the pandemic ends, and everything goes back to 'normal', is there gone be a structural change in the consumer behavior regarding the food service industry?

5. Limitations and future research

A fist limitation of this research is the small number of reviews of the pandemic period. Collecting more data from this period can significantly change the results. The review website selection is also a limitation of this paper. The selection of only TripAdvisor can induce a platform bias. This website focus is on travelling rather than everyday reviews. In this sense, the abrupt reduction on the number of reviews can be explained by the travel ban/restriction amongst states and countries. It is possible that online delivery platforms such as iFood and Uber Eats experienced a 'boom' in the number of reviews. So, future research can study the impact of the COVID-19 pandemic on other websites/platforms. Also, a comparison among different websites can bring different topics representing different consumer perceptions.

Another limitation was the number of cities chosen. While the sample is representative for these cities, it may not represent Brazil as a whole. The four cities are capital of the four states on the southern east Brazilian region, the most developed of the country. Given the continental size of Brazil, as well as the huge cultural differences among the regions and states, a more comprehensive analysis, representing the whole country can bring different results. Finally, future research can combine different datasets in the analysis. For example, economic indicators such as educational level, GDP, human development index, among others, can be used to statistically test differences among the cities. Also, data on the severity of the pandemic per city can be employed to better understand the influence of COVID-19 on the consumer perception of Brazilian food service industry.

Acknowledgement

This research was developed with the support of the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) - Financing Code 001.

References

Balducci, B., & Marinova, D. (2018). Unstructured data in marketing. *Journal of the Academy of Marketing Science*, *46*(4), 557–590. https://doi.org/10.1007/s11747-018-0581-x

Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the Tribes: Using Text for Marketing Insight. *Journal of Marketing*, *84*(1), 1–25. https://doi.org/10.1177/0022242919873106

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*, 993–1022. Retrieved from http://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf

Brotherton, B. (1999). Towards a definitive view of the nature of hospitality and hospitality

management. *International Journal of Contemporary Hospitality Management*, *11*(4), 165–173.
Chheda, S., Duncan, E., & Roggenhofer, S. (2017). Putting customer experience at the heart of next-generation operating models. *McKinsey Quarterly*, *21*, 1–10. Retrieved from https://blog.lemonade.com/2017/01/01/lemonade-sets-new-world-record/?utm_source=instantclaim&utm_medium=facebook&utm_%0Ahttps://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/putting-customer-experience-at-the-heart-of-next-generation-o
Chung, N., & Koo, C. (2015). The use of social media in travel information search. *Telematics and Informatics*, *32*(2), 215–229. https://doi.org/10.1016/j.tele.2014.08.005
Feinerer, I., & Hornik, K. (2019). tm: Text Mining Package. Retrieved from https://cran.r-project.org/package=tmle
Feldman, R. (2013). Techniques and applications for sentiment analysis: The main applications and challenges of one of the hottest research areas in computer science. *Communications of the ACM*, *56*(4), 82–89. https://doi.org/10.1145/2436256.2436274
Fellows, I. (2018). wordcloud: Word Clouds. Retrieved from https://cran.r-project.org/package=wordcloud
Gonzaga, S. (2017). lexiconPT: Lexicons for Portuguese Text Analysis.o Title. Retrieved from https://cran.r-project.org/package=lexiconPT
Grün, B., & Hornik, K. (2011). topicmodels: An R Package for Fitting Topic Models. *Journal of Statistical Software*, *40*(13), 1–30. https://doi.org/10.18637/jss.v040.i13
Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, *59*, 467–483. https://doi.org/10.1016/j.tourman.2016.09.009
Hemmington, N. (2007). From service to experience: Understanding and defining the hospitality business. *Service Industries Journal*, *27*(6), 747–755. https://doi.org/10.1080/02642060701453221
Ju, Y., Back, K. J., Choi, Y., & Lee, J. S. (2019). Exploring Airbnb service quality attributes and their asymmetric effects on customer satisfaction. *International Journal of Hospitality Management*, *77*(November 2017), 342–352. https://doi.org/10.1016/j.ijhm.2018.07.014
Kannan, P. K., & Li, H. "Alice." (2017). Digital marketing: A framework, review and research agenda. *International Journal of Research in Marketing*, *34*(1), 22–45. https://doi.org/10.1016/j.ijresmar.2016.11.006
Khalil, S. (2018). Rcrawler: Web Crawler and Scraper. Retrieved from https://cran.r-project.org/package=Rcrawler
Korfiatis, N., Stamolampros, P., Kourouthanassis, P., & Sagiadinos, V. (2019). Measuring service quality from unstructured data: A topic modeling application on airline passengers' online reviews. *Expert Systems with Applications*, *116*, 472–486. https://doi.org/10.1016/j.eswa.2018.09.037
Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of Marketing*, *80*(6), 69–96. https://doi.org/10.1509/jm.15.0420
Liu, Y., Teichert, T., Rossi, M., Li, H., & Hu, F. (2017). Big data for big insights: Investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews. *Tourism Management*, *59*, 554–563. https://doi.org/10.1016/j.tourman.2016.08.012
Luo, J. M., Vu, H. Q., Li, G., & Law, R. (2020). Topic modelling for theme park online reviews: analysis of Disneyland. *Journal of Travel and Tourism Marketing*, *37*(2), 272–285.

https://doi.org/10.1080/10548408.2020.1740138
Morgan, B. (2017). AI Can Comb Through Your Data to Create More Compelling Customer Experiences. *Harvard Business Review*.
Nagji, N., Venkatesan, R., Trzcinski, R., & Craddock, J. (2018). Automation of Marketing Models. *Darden Business Publishing*, *195386*, 1–9.
Nikita, M. (2020). ldatuning: Tuning of the Latent Dirichlet Allocation Models Parameters. Retrieved from https://cran.r-project.org/package=ldatuning
R Core Team. (2020). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
SEBRAE. (2019). Bars and restaurants: an expanding sector. Retrieved from https://www.sebrae.com.br/sites/PortalSebrae/artigos/bares-e-restaurantes-um-setor-em-expansao,1038d53342603410VgnVCM100000b272010aRCRD
Silge, J, & Robinson, D. (2016). tidytext: Text Mining and Analysis Using Tidy Data Principles in R. https://doi.org/https://doi.org/10.21105/joss.00037
Silge, Julia, & Robinson, D. (2017). *Text Mining with R Text Mining with R Revision History for the First Edition*. Retrieved from http://oreilly.com/catalog/errata.csp?isbn=9781491981658
Statista. (2019). Foodservice as share of food retail sales in Brazil from 2015 to 2019. Retrieved from https://www.statista.com/statistics/726158/market-share-consumer-foodservice-food-industry-brazil/
Statista. (2020). Number and change of coronavirus (COVID-19) cases and deaths among the most impacted countries worldwide as of July 13, 2020. Retrieved from https://www.statista.com/statistics/1105264/coronavirus-covid-19-cases-most-affected-countries-worldwide/
Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, *37*(2), 267–307. https://doi.org/10.1162/COLI_a_00049
Timoshenko, A., & Hauser, J. R. (2019). Identifying customer needs from user-generated content. *Marketing Science*, *38*(1), 1–20. https://doi.org/10.1287/mksc.2018.1123
Wedel, M., & Kannan, P. K. (2016). Marketing analytics for data-rich environments. *Journal of Marketing*, *80*(6), 97–121. https://doi.org/10.1509/jm.15.0413
WHO. (2020a). Coronavirus disease (COVID-19) pandemic. Retrieved from https://www.who.int/emergencies/diseases/novel-coronavirus-2019
WHO. (2020b). WHO Timeline - COVID-19. Retrieved July 16, 2020, from https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19