

Repositório ISCTE-IUL

Deposited in *Repositório ISCTE-IUL*:

2021-01-13

Deposited version:

Accepted Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Fernandes, E., Moro, S., Cortez, P., Batista, F. & Ribeiro, R. (2021). A data-driven approach to measure restaurant performance by combining online reviews with historical sales data. *International Journal of Hospitality Management*. 94

Further information on publisher's website:

[10.1016/j.ijhm.2020.102830](https://doi.org/10.1016/j.ijhm.2020.102830)

Publisher's copyright statement:

This is the peer reviewed version of the following article: Fernandes, E., Moro, S., Cortez, P., Batista, F. & Ribeiro, R. (2021). A data-driven approach to measure restaurant performance by combining online reviews with historical sales data. *International Journal of Hospitality Management*. 94, which has been published in final form at <https://dx.doi.org/10.1016/j.ijhm.2020.102830>. 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.

A data-driven approach to measure restaurant performance by combining online reviews with historical sales data

ABSTRACT

Restaurant management requires customer responsiveness to deal with increasingly higher expectations and market competitiveness. This study proposes an approach to simplify the decision-making process of restaurant managers by combining both live social media customer feedback and historical sales data in a sales forecast model (based on TripAdvisor data and the Bass model).

Our approach was validated with internal and external (i.e., online reviews) data gathered from six restaurants. The collected data was processed using data analytics for developing a dashboard that provides value for restaurateurs by taking advantage of online reviews and sales forecast. Such dashboard was evaluated by restaurant management experts, which provided positive feedback, highlighting in particular the time saved in the decision-making process.

Keywords: restaurant management; business performance; customer relationship management; online review; text mining; data analytics.

1. Introduction

Revenue management enables to optimize businesses toward revenue maximization by understanding consumer behavior and implementing a product and price strategy accordingly (Kimes, 1999). The market dynamics demands for suitable and constant business performance measuring to monitor the business and revenue progress (Cross et al., 2009). Within hospitality, the number of customers that arrive at a given hospitality unit is a key input to forecast revenue (Weatherford & Kimes, 2003). Restaurant researchers and practitioners devised models based on the number of customers per time unit (Heo, 2017). The Bass model (BM) has been widely adopted to forecast demand by predicting the number of new customers in the forthcoming period (Mahajan et al., 1991). Specifically in the hospitality literature, the BM has been applied to hotel management (e.g., Pimpão et al., 2016). Furthermore, restaurant customers, whether they are innovators or imitators (according to the BM), may lead to successive increases in the number of new customers, which can be modeled through the BM (Sultan et al., 1990). However, within our knowledge, there are no research studies that adopted the BM to restaurant management.

Consumer-generated content has been steadily increasing since the paradigm of the Web 2.0 brought us social media online platforms where all the contents are created by users (Hennig-Thurau et al., 2004; Wang & Rodgers, 2011). As users take interchangeably the role of producers and consumers of information in social media, electronic word-of-mouth comes into play to develop communication networks that influence users in their decision processes (Costa et al., 2019). One popular format of communication in social media in hospitality is the online review (OR), which enables guests to write their opinions about the units and attractions they visited (Moro et al., 2019a). ORs were found to be critical for understanding hotel success (Xiang et al., 2015). As for restaurants, Jalilvand et al. (Jalilvand et al., 2017) studied the factors influencing electronic word-of-mouth, arguing that customer satisfaction influences business

performance. Yet, they did not analyze such impact on sales. The recent study by Fernández-Miguélez et al. (Fernández-Miguélez et al., 2020) established a relation between ORs and financial performance by analyzing macro-information based on firms financial reports.

There is more evidence supporting a relation between ORs and business performance in the OR body of knowledge within hospitality. Kwok et al. (2017) analyzed a total of 67 articles focused in OR and published between 2000 and 2015. They conclude that ORs effectively enable to better understand business performance and outcomes. However, most of the studies were related to hotels, and only eight of them were about restaurants. From those eight studies, none has devoted attention to measuring the customer inflow based on OR. Most of them are related to management response strategies, the usefulness of OR, and a few specific themes such as group buying. Thus, Kwok et al. (Kwok et al., 2017) conclusion of ORs influence in business performance is based in hotel studies (W. Kim et al., 2015; K. L. Xie et al., 2014).

Fan et al. (Fan et al., 2017) propose a combination between the BM (F. Bass, 1969) and sentiment analysis that results in a novel method for sales forecast, the Bass Emotion model (BEM). The BM, known as a diffusion model, models the process of product adoption by customers. Moreover, previous studies showed that the diffusion process is affected by Word-of-mouth (WOM) (Sultan et al., 1990) which supports for the combination of the BM and information from ORs. Restaurateurs usually forecast demand by using a methodology from a range of analytical models available on existing literature (Lasek et al., 2016). Each model has a set of input variables as predictors. However, we did not find a model to forecast restaurant sales that uses ORs as input. In this study, we take advantage of the BEM and adapt it to suit the purpose of forecasting sales of six restaurants based on the Algarve region of Portugal. To facilitate the evaluation of the BEM results, we developed a dashboard to present the information for decision support, including a new key performance indicator (KPI) computed based on restaurant performance and customer satisfaction. Therefore, the resulting

contribution from our study helps restaurateurs to manage their units by forecasting sales and monitoring the direct impact of daily operations on customer satisfaction and sales.

This article is structured as follows. Next section is devoted to reviewing related literature. The methodology section presents the research design and the proposed data-driven approach. In the following section, the results are described and analyzed, including a critical discussion. Finally, the conclusions section presents the main contributions and implications from the study.

2. Literature review

2.1. Online reviews analysis to improve performance in the foodservice industry

Online reviews (ORs) have become the mainstream medium for sharing feedback about products and services (Neirotti et al., 2016). The hospitality industry has been at the forefront of developing and adopting ORs platforms, including the renowned brands TripAdvisor and Yelp (Guerreiro & Moro, 2017). The vast majority of published studies within hospitality based on ORs are focused on accommodation services (e.g., hotels: Moro et al., 2020). Nevertheless, the uniqueness of the foodservice industry combined with the increasing relevance of online reviews has also triggered research to understand the perspectives of restaurants customers, helping to shape practitioners' strategies (W. G. Kim et al., 2016).

The richness of a textual review makes it suitable to analyze the customer's point of view. However, the challenge of dealing with usually large numbers of reviews as well as the unstructured nature of text requires an automated approach for an efficient analysis (Aggarwal & Zhai, 2012). Text mining (TM) approaches enable to extract useful knowledge from a corpus of documents (or reviews), containing unstructured text, through techniques that compute features and their corresponding weights (e.g., word frequency) from text (Miner et al., 2012). A key component of TM is natural language processing (NLP) as it comprehends a set of

different tasks devoted to extract a meaningful representation of text, such as part-of-speech tagging (i.e., by identifying key components such as nouns, verbs, and adjectives from sentences) and sentiment analysis (Kao et al., 2007). Sentiment analysis (SA) can be considered a regression problem by assigning a sentiment score that reflects the sentiment polarity (i.e., 0 representing a neutral sentiment, while positive/negative numbers represent the polarity – through the sign – and the intensity – through the absolute value) to a given sentence of the text (Batista and Ribeiro, 2013).

Online reviews can deliver important information to improve hospitality units' performance (Xiang et al., 2015) by helping to understand which factors influence customer opinion (Berezina et al., 2015; Han et al., 2016; Moro et al., 2019b; Nitiwanakul, 2014; O'Connor, 2010; Zhang & Verma, 2017). Most of the studies are dedicated to the relation between traditional attributes (Gan et al., 2016; Golani et al., 2017). Yet, within foodservice research, literature is rather scarce in taking advantage of ORs to assess restaurant performance using data-driven approaches based on sales data (S. Kim & Kim, 2016). Table 1 details three studies within foodservice specifically adopting OR and sales performance data, with the older from 2010. Thus, while the theme is not novelty, the lack of access to corporate sales data seems to limit the number of publicly available studies.

Table 1

Foodservice studies adopting both ORs and sales performance data

Analytical Method, Research Goal and Major Findings	Reference
Analytical Method: Ordinary Least Squares (OLS) Research Goal: To test the impact of WOM on corporate financial performance. Major Findings: Corporations with a high value of WOM in their establishments will also have higher profitability levels. The study proved that the positive effect of WOM on the financial performance of establishments also impacts the financial performance of corporations.	(Fernández-Miguélez et al., 2020)
Analytical Method: Hierarchical multiple regressions Research Goal: To identify the new or unexplored determinants of the restaurant. Major Findings: The number of ORs and the customer overall rating have a positive impact on restaurant performance. Both variables have a greater impact on promoting net sales in restaurants with an excellence certificate.	(Kim et al., 2016)
Analytical Method: Regression model Research Goal: To analyze the differences between online user-created reviews and editor-created reviews. Major Findings: Consumer-reviews and their volume can increase the online popularity of a restaurant. Editor-created reviews have a negative impact on the consumer's intention to visit the restaurant.	(Zhang et al., 2010)

2.2. Restaurant performance and business KPI's

Kimes (1999) defines Restaurant Revenue Management (RRM) as selling the right seat to the right customer at the right price and for the right duration (Kimes, 1999). The development of models to forecast demand can help in guiding managers towards successful RRM strategies (Lasek et al., 2016). To improve the decision making process, efficiency measures, also known as key performance indicators (KPI), are computed based on several inputs, including the results from sales forecast models (S. Kim & Kim, 2016). To increase interpretability and facilitate reading toward improved decision support, usually KPIs are shown in visually appealing dashboards (Pestana et al., 2020). Nevertheless, a RRM strategy is usually complex and requires encompassing a myriad of inputs (W. G. Kim et al., 2016)

According to the KPI Institute (The KPI Institute, 2016), the most frequent KPIs used in restaurants can be grouped as follows: customer feedback, occupancy, service measurement, revenue, cost management, human resources, and quality compliance. For instance, the

proportion of positive feedback from guests is a customer feedback KPI usually presented separately and unrelated from the number of tables served or the revenue per available seat hour (RevPASH). Similarly, KPIs are not usually viewed in combination with the expected demand (The KPI Institute, 2016).

There are several types of restaurants according to the customer target. Knutson et al. (Knutson et al., 2008) identified three main segments: quick service, casual/theme, and fine dining. The RRM strategy should be adjusted to each target (Thompson, 2010), which implies the KPIs need to be interpreted within the target segment. Subsegments may also be defined within the three above mentioned segments. For example, casual restaurants may be standard or premium, depending on the service level (Saad et al., 2020), while fine dining may become luxury if there is an emphasis on symbolic and conspicuous values associated with a more relaxed environment instead of being mostly focused on the food quality (Yang & Mattila, 2016).

Usually, each KPI is adopted as a base for defining strategies for Kitchen Management, Front of House & Restaurant Management, Bar & Cellar Management, Sales & Marketing Management, or Finance & Administration Management. However, the existing body of knowledge lacks in studies that report any KPI combining customer feedback information (such as ORs information) and business performance metrics to understand the factors that drive business towards a successful RRM.

2.3. Sales forecast using online reviews

Sales forecasting is essential in business management (Chern et al., 2015; Lasek et al., 2016). Moreover, ORs become a strong influence factor on customer purchase decision (Fernández-Miguélez et al., 2020; Kwok et al., 2017). In fact, recent studies show a relation between ORs and sales by showing that it is possible to forecast product sales through ORs. For instance, in the movie domain, Yu *et al.* (2012) proposed an Autoregressive Sentiment and

Quality Aware model to predict sales by using the sentiment expressed in the reviews (Yu et al., 2012). In the retail domain, Chern *et al.* (2015) presented a sales forecast model by combining ORs and a linear regression model. Furthermore, in the automotive industry, a sales forecast model combining BM and sentiment analysis was proposed by Fan *et al.*, (2017). The particular research was cited 120 times (as of 11 July 2020) in the Scopus database, showing that the approach of Fan *et al.* (2017) is applicable to other contexts, including the hospitality domain by considering hotel ORs (e.g., Aakash & Gupta Aggarwal, 2020).

The BM is a kind of diffusion model that forecasts the adoption of products by considering the influence of external publicity or promotion effect, as well as the diffusion effect introduced by ORs (F. Bass, 1969; F. M. Bass, 2004). The model predicts the number of customers from the increases in the number of adopters that can be classified into innovators (consumers that buy under external influence) or imitators (consumers that buy under internal influence) (Lai, 2017; Sultan et al., 1990).

In our sample of restaurants, the sales curve is remarkably close to Roger's adoption curve (Rogers, 2003), as well as the BM curve. Typically, the year starts with a low number of customers, and as time progresses, the number of customers increase until a peak, and then the number of customers starts decreasing. Furthermore, within our knowledge, there is no research study that applied the BM to the specific and important restaurant sector. Thus, in this paper, we aim to present an approach that improves the original BM in the specific context of restaurant domain. In the BM, the predicted number of customers who have adopted the product in the period t , is computed as follows:

$$S(t) = m \frac{\frac{(p+q)^2}{p} e^{-(p+q)t}}{\left(1 + \left(\frac{q}{p}\right) e^{-(p+q)t}\right)^2}$$

where $S(t)$ is explained by three parameters: p , the coefficient of innovation, q , the coefficient of imitation and m is the potential market, which can be the total number of ultimate adopters (F. Bass, 1969). The cumulative sales at time t is given as (Lai, 2017):

$$Y(t) = m \frac{1 - e^{-(p+q)t}}{1 + \left(\frac{q}{p}\right)e^{-(p+q)t}}$$

Both can be presented by the following equations:

$$S(t) = pm + (q - p)Y(t) - \left(\frac{q}{m}\right)[Y(t)]^2, Y(t) = m \left[\frac{1 - e^{-(p+q)t}}{1 + \left[\frac{q}{p}\right]e^{-(p+q)t}} \right]$$

The time of peak adoptions is achieved when $t = \frac{1}{p+q} \ln \frac{p}{q}$ and the number of adopters at the peak time is given by $s = \frac{m}{4q}(p + q)^2$.

The BM became widely used in theoretical research and practical applications (Mahajan et al., 1991) as result of its low predictive error (Zhang et al., 2020). In 2004, Bass seminal paper was selected as one of the ten most frequently cited papers in the 50-year history of the Management Science domain (Massiani & Gohs, 2015). However, the model has advantages and disadvantages. It can explain in a simple way the existence of an empirical generalization. Furthermore, the basic assumptions and the calculated parameters provide an intuitive explanation. It also allows to analyze the impact of innovators and imitators, and understand the best time to launch a new innovation in the product or service. The model only uses historical data and its main parameters are assumed to be constant throughout the diffusion process, although in practice they often vary (Fan et al., 2017; Zhang et al., 2020). Thus, in order to improve the model, studies presented different approaches to calculate p , q and/or m (Chern et al., 2015; Fan et al., 2017; Yu et al., 2012). Recently, Zhang *et al.* (2020) presented a

study to improve the predictive power of the BM by using ORs, search traffic data, and macroeconomic data to calculate m , p , and q . Our approach is presented in Section 4.3.

3. Methodology

3.1. Research Design

We adopted and adjusted the widely used cross industry standard process for data mining (CRISP-DM) methodology to develop our proposal (Chapman et al., 2000) (Fig. 1). The adapted CRISP-DM consists in cyclic steps towards improving BEM forecast while providing an intuitive dashboard to enable restaurateurs to assess the relevance of the results and hence validate the model's usefulness in the foodservice industry to the decision-making process. All the experiments were developed using the open source R statistical tool, thus benefiting from an enthusiastic community of supporters contributing with packages for a myriad of data analysis tasks (Cortez, 2014). Each grey box in Fig. 1 denotes a step in the process, with the adopted R packages highlighted in each arrow.

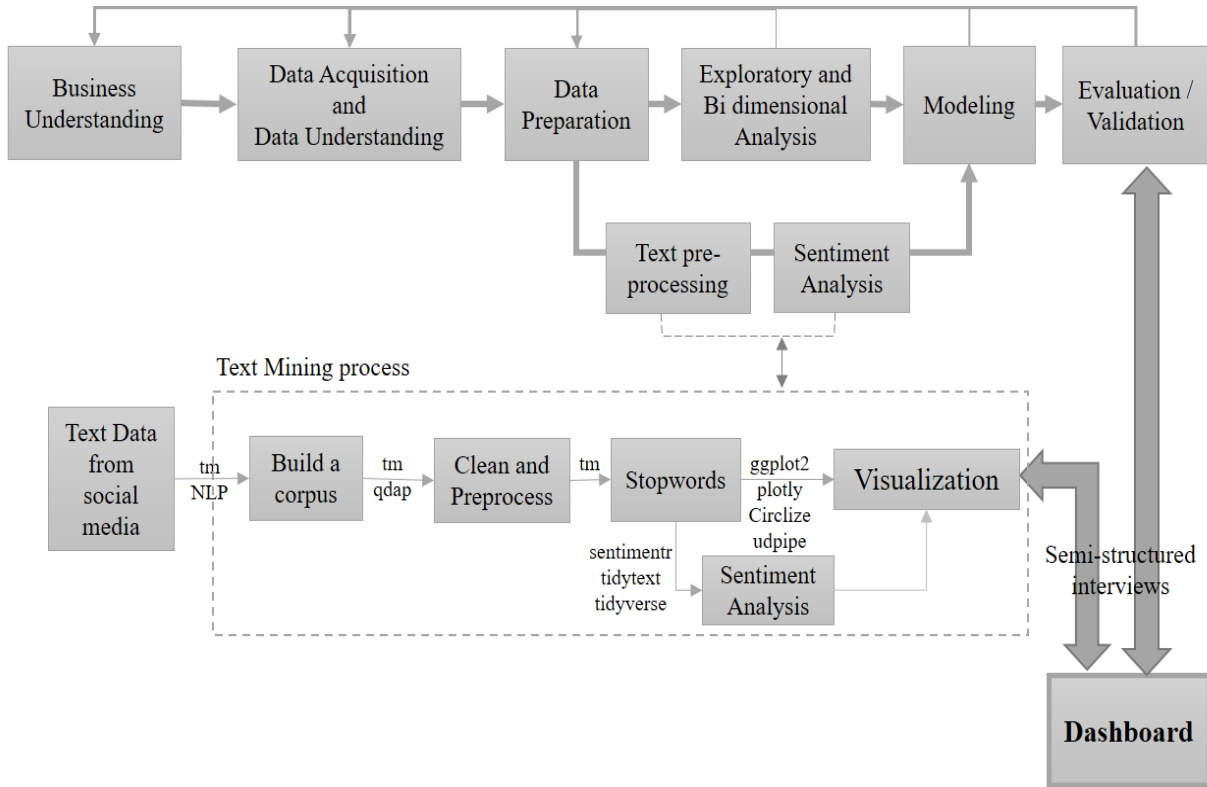


Fig. 1 Schematic of the adopted methodology.

The entire process is detailed in the following sections. First, we start by a descriptive data analysis followed by a bi-dimensional analysis for data understanding. Then, we detail the TM process implemented to extract the factors mentioned in the ORs and to calculate the sentiment score. In the results section, we present TM results, such as the co-occurrences between nouns (main elements) and adjectives (qualifiers), followed by the results of the proposed KPI and the BEM sales forecast. Then, the information is consolidated into a dashboard for easier validation by experts towards improved decision support. Thus, the proposed dashboard provides the sales forecast that reflects the impact of customer satisfaction on restaurant performance.

3.2. Data Understanding and Data Analysis

This study adopted two data sources from six restaurants: the revenue management information system, and the TripAdvisor. For both, we stored the data values on a daily basis, encompassing a period ranging from January 2015 to September 2017. A total of 1,220 reviews

publicly available and written in English were retrieved. More than half of the reviews were written from customers of the United Kingdom (51%) followed by customers from Ireland (11%) and Portugal (7%). The extracted features were the following: rating or review score (a 5-point scale from 1-terrible to 5-excellent), review date, the title (represents an experience summary), the review text (a written account of the experience), the customer location (city and country), and the total number of reviews published by the customer on the TripAdvisor website. From the revenue system, we collected the number of customers per day, by restaurant, and the average amount spent (Yield) per customer.

For confidentiality reasons, we only mention that our restaurants are based in the Algarve region in Portugal. Fictional names are used, i.e., restaurants are labeled from Restaurant A to F. We classified those restaurants by restaurant segment (Knutson et al., 2008). Restaurants A and B are “luxury” restaurants and C is a “fine dining” restaurant (Table 2) (Knutson et al., 2008). Restaurants D, E, and F are classified as premium casual restaurants (Knutson et al., 2008). We conducted a descriptive and bi-dimensional analysis of the business and TripAdvisor KPIs.

All restaurants have more than one hundred daily customers except restaurant C, with an average of 28 customers per day. Accordingly, restaurant C has less reviews (70 reviews) although exhibiting the highest average rating of 4.7. Customers who wrote a review about restaurant C are experienced in sharing online experiences because they have an average of 64 contributions on TripAdvisor. Restaurant A has an average of 113 customers per day, it has more reviews (352) and a lower average rating (4.0). Restaurants B and F have 232 and 272 reviews respectively, with an average rating of 4.2 and 4.1. Finally, restaurants D and E have less than 200 reviews each (180 and 114), an overall rating of 4.1 and 4.4, respectively.

3.3. Correlation analysis among variables

In this section, we analyzed correlations between business KPIs and TripAdvisor KPIs (Table 2). We calculate the Spearman Coefficient (SC) to measure relations of strength between number of customers per day, Yield, and Rating. This coefficient measures the monotone association between variables. It is used when one or both variables are ordinal (Hauke & Kossowski, 2011; Mukaka, 2012; Spearman, 1904).

There is a moderately negative relation between business KPIs and Rating (Table 2). To assess the reason for such result, restaurant managers were interviewed in the evaluation step of our methodology (Fig. 1). When faced with the results, they argued that such results are understandable because the increasing number of customers directly reflects the waiting time which leads to a reduction in service quality that decreases customer satisfaction (Hwang, 2008). This happens for restaurants A, B, D, and E. Furthermore, both luxury restaurants (A and B) combine characteristics that influence customers' length of stay, such as the room configuration and flexibility of seating that require effective table management (Hwang, 2008). Restaurants D and E are premium casual restaurants, frequented by customers with opposite goals: families with kids that enjoy the playground space, and senior golfers who want to be in a quiet ambiance.

The Yield increase induces a Rating decrease because higher prices are related to special events where customers have higher expectations. As they seek a memorable experience, there is a narrow 'tolerance zone' (Golani et al., 2017). A delighted customer increases the positive ORs but the smallest detail drops dramatically the satisfaction and consequently the granted Rating score (Golani et al., 2017; McGuire, 2016).

Table 2

Descriptive and Bi-dimensional analysis

Restaurant		A	B	C	D	E	F
Restaurant type		Luxury		Fine dining	Premium casual		Premium casual close to the beach
Business KPIs							
No. of customers per day (average)		113	124	28	127	107	134
No. of Reviews		352	232	70	180	114	272
TripAdvisor KPIs							
Rating (average)		4	4.2	4.7	4.1	4.4	4.1
User Experience on TripAdvisor (Avg. reviews)		41	51	64	49	58	55
Spearman rho Coefficient (SC) (**)							
No. of Customers per day	Rating	-0.43	-0.52	-0.36	-0.7	-0.52	-0.25
	p-value	0.000*	0.000*	0.002*	0.000*	0.000*	0.000*
Yield	Rating	-0.26	-0.41	-0.51	-0.21	-0.6	-0.27
	p-value	0.000*	0.000*	0.000*	0.005*	0.000*	0.000*

(*) $p \leq 0.05$ the coefficient is highly significant, (**) $p \leq 0.10$ the coefficient is marginally significant (Filho et al., 2013);
 (**) Values between 0.4 and 0.79 indicates a moderately strong correlation (Swinscow & Campbell, 2002)

3.4. Text Mining and Sentiment Analysis

Restaurateurs can optimize their business decisions and consequently increase customer satisfaction by knowing customers' opinions. In order to unveil the most mentioned factors in the ORs according to the level of customer satisfaction, we conducted a text mining analysis including a sentiment score computation. For corpus handling and text preprocessing we used the *tm* (Feinerer & Horik, 2018), the *NLP* (Hornik & Hornik, 2018), and the *qdap* packages (Jovi et al., 2015) from the R statistical tool. Preprocessing involved punctuation, numbers, and white spaces removal, as well as to remove common English words, such as “the” or “by”, with little semantical value (we used the list of words considered in the function *stopwords* from the *tm* package). Stop words removal reduces the number of words in the document and increases the effectiveness and efficiency of text processing (Irfan et al., 2019). Furthermore, another list

of words, such as restaurant brands or abbreviations, were removed. We also standardized the English words to British (UK) English by replacing American English words by the equivalent ones from the British dictionary. Moreover, to retrieve collocations and co-occurrences, i.e., words that occur together and words that are followed by another, the package *udpipe* was used. We aimed to find combinations of adjectives and nouns that contribute to customer satisfaction. We computed the sentiment score of title and review through the *sentimentR* package to measure levels of customer satisfaction. *SentimentR* was designed to quickly calculate text polarity sentiment (TPS) at the sentence level and optionally aggregate by rows or grouping variable(s) (True, 2018). In fact, we aimed to find factors related to high and low TPS levels. For instance, we can understand the frequency between the words “great restaurant food” to see which paired combination contributes more to high TPS. As explained in section 2.3, the sentiment score is then used as q coefficient to reflect the customers' preferences into the BEM model, as proposed by Fan et al. (Fan et al., 2017).

4. Results and discussion

Given the exploratory nature of this research, we structured this section as follows. First, we present a descriptive analysis and factors of customer satisfaction or dissatisfaction based on the adopted TM techniques. Second, we describe a new performance indicator and the proposed sales forecast model. Finally, we combine all the information into a dashboard to facilitate the evaluation of the management information, including the sales forecast, by a panel of restaurant management experts.

4.1. TPS analysis and factors of customer satisfaction or dissatisfaction

Table 3 presents statistics for both Rating and TPS, as well as the correlation between both. As expected, restaurants with higher ratings have ORs with strong positive sentiments. Restaurant E has a high TPS (0.90) followed by restaurant C (0.89). These results are further

corroborated by the positive and high correlation (0.91) between Rating and TPS average. We further note that the variability (standard deviation – std) of the Rating and TPS scores is higher in restaurant A because it is the restaurant with the highest variety of dining events. Furthermore, the variability of the TPS in restaurants C and D is smaller because they have consistent service quality and an efficient table management strategy that, as argued in (Hwang, 2008), prevents customer dissatisfaction.

Table 3

Restaurant average ratings, average TPS, and Pearson coefficient between TPS and rating.

Restaurant	Rating (average)	TPS of Title + Review (average)	TPS vs Rating (Pearson coefficient)
Restaurant A	4.0 (std= 1.20)	0.63 (std= 0.57)	0.55
Restaurant B	4.2 (std=1.15)	0.68 (std=0.47)	0.41
Restaurant C	4.7 (std=0.48)	0.89 (std=0.43)	0.20
Restaurant D	4.1 (std=1.17)	0.68 (std=0.43)	0.38
Restaurant E	4.4 (std=1.03)	0.90 (std=0.50)	0.41
Restaurant F	4.1 (std=1.15)	0.64 (std=0.47)	0.38
Pearson coefficient between average rating and average text sentiment (TPS): 0.91			

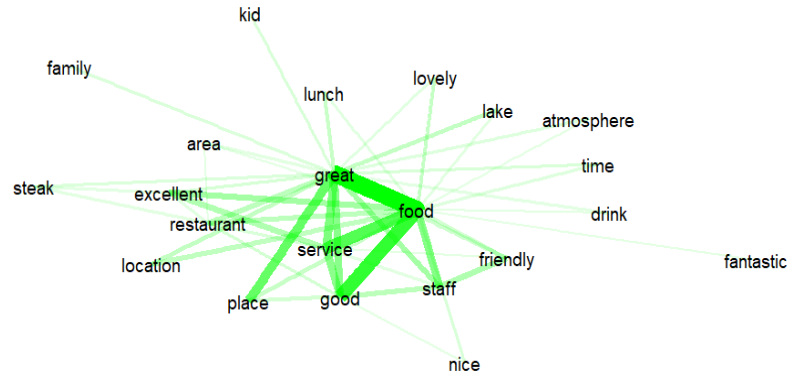
In addition, the Pearson coefficient between TPS and Rating is stronger in restaurant A, which indicates that, when TPS increases, the Rating given by the customer will also be higher. We also computed the average TPS by country. Customers with lower TPS were from New Zealand, Canada, and South Africa. Customers from the United Kingdom and Ireland were the most frequent restaurant visitors (representing 80% of the customers and 62% of the TripAdvisor reviewers), but they were not the most satisfied (average TPS of 0.66 and 0.77 respectively). Such information is helpful to define managerial and marketing strategies, for example, to address the issues raised by the most frequent customers.

Both negative and positive reviews are potentially important for restaurateurs daily decisions (Phillips et al., 2017). However, since all restaurants have an average Rating greater than four (Table 3), the most frequent words in ORs concurrently reflect positive sentiments. To confirm this, we used the *tidytext* package (Silge & Robinson, 2017) to assess how the

different sentiments are represented across the reviews. We found a total of 4,466 positive and 738 negative words, thus revealing an overall stronger positive customer opinion. Nevertheless, customer feedback evolves through time. For instance, we have identified that the average TPS by month presents a decrease between June and October, which occurred while there was an increase in the number of customers.

To facilitate the evaluation of the dynamics of the information, the proposed dashboard includes a temporal dimension for showing relevant Rating and TPS indicators (e.g., last day, week, or month). The dashboard also includes a visual word network representation (shown in Fig. 1), which is adapted to a time period and that is related to the most frequent words by Rating and also the combination of adjectives and nouns.

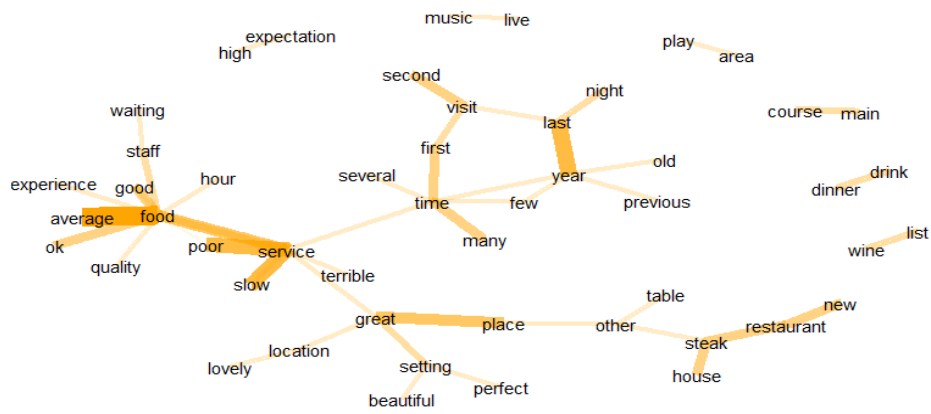
As described in section 3.4, we include a co-occurrence analysis to assess the main links between the words mentioned in the reviews by customers. All nouns and adjectives were used. This enables to assess, for example, if a “restaurant” considered as “excellent” is more linked to “food” or to “service”. Fig. 2 (a) represents the network with the most frequent word combinations in all the reviews. According to current body of knowledge, negative ORs can have a dramatic impact on the business (McGuire, 2016; Sparks et al., 2016; K. Xie et al., 2017). Thus, negative factors are presented in the network of Fig. 2 (b), which represents the most frequent word combinations in the reviews with less than four in the Rating scores. Finally, the network in Fig.2 (c) represents the words following other terms in negative reviews.



(a)



(b)



(c)

Fig. 2 Word network representations: (a) co-occurrences within sentence in all the reviews (nouns and adjectives); (b) co-occurrences in the less positive reviews (Rating between 1 and 3); (c) words following one another in less positive reviews.

For instance, “poor service” or “average food” are two frequent combinations in the ORs. As previous studies showed, in a fine-dining restaurant, food, physical environment, service and price are factors mentioned in ORs to describe the restaurant experience (Boo, 2017; Hsu et al., 2018; Nitiwanakul, 2014; Spyridou, 2017). The presented networks can complement the decision-making process by providing a roadmap for improvements.

4.2. KPI and restaurant performance

Restaurateurs monitor daily KPIs to analyze business performance and make decisions in order to maximize revenue. We propose a new performance indicator (KPI_{Sent}) that combines business performance and customer perspective. This KPI consists of the sum of four ratios:

- the average Rating in the period I (that varies between 0.2 and 1), by the maximum Rating possible;
- the average TPS of the title and the review in period i by the maximum TPS ever;
- the number of customers in period i by the maximum seats available for the season, which means in summer some restaurants have more available tables than winter; and
- the average Yield in period i by the maximum Yield ever.

$$KPI_{Sent}(i) = 1 + \frac{Avg.rating_i}{Max(rating)} + \frac{Avg.Sent_i}{Max(SentTitle)} + \frac{Nr customers_i}{Max Nr Customers} + \frac{Yield_i}{Max(Yield)}$$

$$\text{where } \frac{1}{5} \leq \frac{Avg.rating}{Max(rating)} \leq 1, \quad -1 \leq \frac{Avg.Sent}{Max(SentTitle)} \leq 1, \quad 0 \leq \frac{Nr customers}{Max Nr Customers} \leq 1,$$

$$0 \leq \frac{Yield}{Max(Yield)} \leq 1,$$

KPI_{Sent} varies between 0.2 (level ‘Poor’) and 5 (level ‘Excellent’) and can be calculated for any period. Ideally, the restaurateur should have high values of KPI_{Sent} , which means a

high number of customers that pay high values for the experience, and that express high levels of satisfaction. When the KPI_{Sent} presents low values, the restaurateur needs to see in the dashboard if the origin was a low number of customers, a low value of yield, or low levels of satisfaction. For instance, if KPI_{Sent} is 2.5, in a period with a high number of customers and high average Yield, it means that customers presented dissatisfaction factors in ORs and the dashboard indicates those factors. This KPI is another informative tool that can be used by managers in combination with the co-occurrence network to assess the units' performance.

4.3. Forecast sales using online reviews and historical sales data

To forecast the monthly number of customers from January to September 2017, we used the BM. Thus, the imitation coefficient q is the average TPS. In fact, TPS values are important for prediction because customers who gave the same Rating can have different values of TPS (Fan et al., 2017). Furthermore, the external influence (p) can be measured by inspecting the relative search interest (RSI) on Google Trends, i.e., p is the probability of purchase of a candidate customer, in a given period, that can be estimated by RSI (Zhang et al., 2020). However, the potential customers are affected by mass media that promotes resort brands, or golf camps brands, where the restaurant belongs. Thus, customers do not search the restaurant name and in some months, RSI is null. In the period under study, restaurants A to D present average values of RSI between 0.1 and 0.3, and restaurant E and F does not appear on Google trends. Nonetheless, this model is more affected by WoM than by innovation (Sultan et al., 1990). Thus, we keep p constant as recommended by Sultan et al. (1990), ($p = 0.03$). Moreover, instead of keeping m constant, we calculate the potential market at t month by considering the imitation coefficient of the previous month and the respective sales of the same months in previous years.

$$Average(S_{t-12}, S_{t-24}) = \frac{m_t}{4q_{t-1}}(p + q_{t-1})^2 \Leftrightarrow m_t = \frac{Average(S_{t-12}, S_{t-24})4q_{t-1}}{(p + q_{t-1})^2}$$

Forecasted sales were computed as follow:

$$S_t = pm_t + (q_{t-1} - p)Y_{t-1} - \left(\frac{q_{t-1}}{m_t}\right)[Y_{t-1}]^2 = \left(p + \frac{q_{t-1}Y_{t-1}}{m_t}\right)(m_t - Y_{t-1})$$

In order to measure the model performance (evaluation step of the knowledge discovery process, see Fig. 1) we decided to use the decomposition model (Yaffee & McGee, 2000) and compare results. For instance, decomposition models are used to define trends and seasonal factors such as holiday effects or agricultural factors in time series. The basic decomposition models can have two structures: additive or multiplicative. In this context, we used an additive model (AM) because the seasonal variation is relatively constant over time and the number of customers in each restaurant has a maximum value. Furthermore, AM is widely applicable and gives a way of decomposing a time series into simple time series. Moreover, it is flexible and constructive method (Quan & Cai, 2009).

Thus, we decompose our time series into seasonality, trend-cycle, and residual effect. First, seasonality is the fluctuation that occurs for each period. In the restaurants under study, there is a regular annual variation of customer's number. Second, trend implies a linear increase or decrease in the time series over a period. Furthermore, the trend can be deterministic or stochastic. Moreover, cyclical fluctuation is the trend variation that results from economic cycles. In fact, business cycles and the tourism demand cycle is cointegrated (Croes & Ridderstaat, 2017). However, the period under analysis benefits of an European financial stability created by the policies implemented as a response to the 2007-8 financial crisis (Maggs, 2020). Moreover, as reported by the World Tourism Organization, between 2010 to 2017, there was a sustained growth in international tourist arrival (UNWTO, 2018). Therefore, in our model, we keep the trend and cyclical fluctuation together in one component, trend-cycle (Yaffee & McGee, 2000). Finally, the residual effect is the random error effect. The overall model is designed as:

$$\text{Number of customers} = \text{Seasonality} + \text{Trend_Cycle} + \text{Residual effect}$$

For each restaurant, the AM was calculated to forecast monthly values from the year of 2017. We smoothed the time series by using a centered moving average, in which we used values from before and after the current time. Then, the trend-cycle was estimated by using linear regression over the smoothed time series. The next step was to obtain a seasonal component by subtracting the trend-cycle estimated from the series. The forecasted values were calculated using the de-trended series and the seasonal. Finally, we calculated the average of the residual effect:

$$\text{Residual effect } (i) = \text{series}(i) - \text{trend}(i) - \text{seasonality}(i)$$

The results from the AM and the BEM were compared. To measure the fit precision we used the root mean squared error (RMSE) and the coefficient of determination (R^2) (Fan et al., 2017; Yaffee & McGee, 2000). To measure the accuracy of the model we used the mean arctangent absolute percentage error (MAAPE) (S. Kim & Kim, 2016) because the mean absolute percentage error (MAPE) is distorted by outliers.

$$\text{MAAPE} = \frac{1}{n} \sum_{i=1}^n \arctan \left(\left| \frac{y_i - \hat{y}_i}{y_i} \right| \right)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}$$

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

As presented in Table 4, the BEM presents: better values of MAAPE in all the restaurants except restaurant F; better values of RMSE in restaurants A, C and D; and better values of R^2

in restaurant A. Furthermore, the BEM achieves values of R^2 greater than 0.6, which also indicates an acceptable level of model performance.

Table 4

Obtained results for the Bass Emotion (BEM) and Additive (AM) models (best values are highlighted using a gray background color)

Restaurant	MAAPE		RMSE		R^2	
	BEM	AM	BEM	AM	BEM	AM
A	0.36	0.51	490	1071	0.92	0.88
B	0.36	0.51	557	550	0.83	0.92
C	0.34	0.47	52	67	0.68	0.9
D	0.44	0.65	690	1193	0.71	0.76
E	0.28	0.36	482	325	0.83	0.85
F	0.39	0.3	1031	592	0.76	0.95

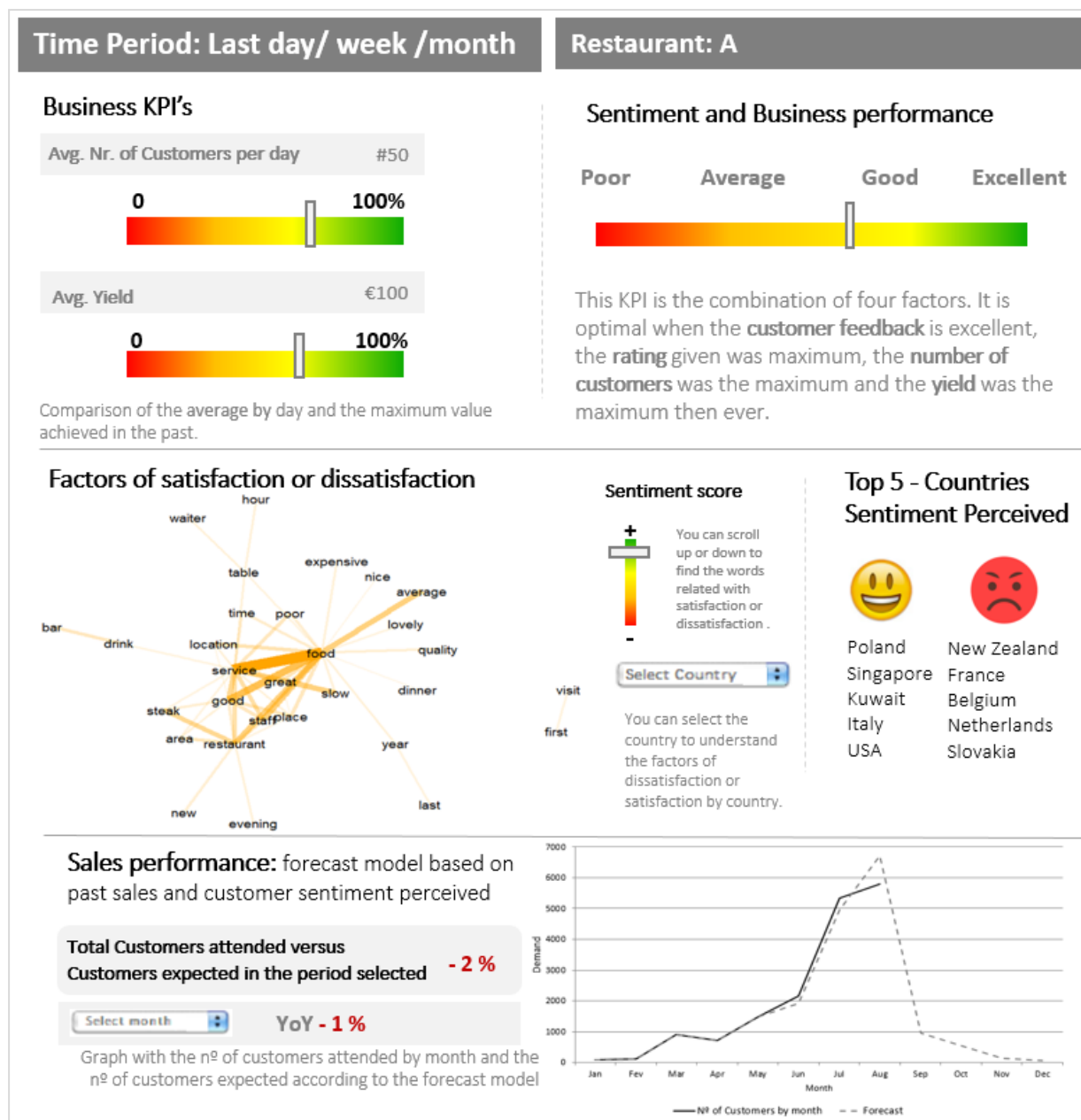
4.4. Evaluation of the information usefulness

In this section, we evaluate the usefulness of the provided information through a dashboard that visually shows the impact of ORs into business performance and highlights factors to improve or promote to maximize sales (Fig. 3). By incorporating the innovative BEM forecast model, this visually appealing display of information can be used by managers (Wexler et al., 2017) through an overview of the past, current, and future performance. In fact, while restaurateurs already have customer information from social networks and business performance KPI's into the same system, the combination of a forecast model with social media feedback has no precedent in the foodservice literature.

The dashboard has three main elements for decision support. First, business performance indicators, such as the average number of customers per day and the average Yield for the period. As well as, $KPI_{Sent}(i)$ that shows the relationship between business performance and customer satisfaction. For instance, if $KPI_{Sent}(i)$ is less than excellent the restaurateur can see in the dashboard if it is caused by a low number of customers and/or Yield, if not, he can explore information from ORs. Then, factors of satisfaction or dissatisfaction are presented

in a visual word network. The restaurateur can see the frequent combinations between words and adjectives pointed by customers by the level of satisfaction. The goal is to prioritize actions to improve service standards or promote attractive elements to, for example, enhance marketing actions. Furthermore, this section can also show the level of customer satisfaction by nationality. Finally, the dashboard presents the number of customers expected as a result from the forecast, and the comparison between real sales and budget. We followed the best practices in dashboard design to facilitate the evaluation of the information provided (Few, 2006; Pestana et al., 2018; Wexler et al., 2017).

The evaluation of the dashboard was made in three steps: first, a hospitality expert was interviewed and provided some pros, cons, and improvements that were applied. Then, we asked a designer to analyze the dashboard. Finally, we interviewed a restaurant manager to understand the applicability and usability of the dashboard. The feedback was positive. The manager argues that the dashboard helped to daily monitor three important business issues: the disappointing factors to improve; the balance between restaurant occupation, revenue, and TripAdvisor feedback; and, finally, the expected impact on sales after an increase of customer's satisfaction. The latter point remarked by the manager highlights the usefulness of the information provided by the sales forecast BEM model.



models available on existing literature (Lasek et al., 2016). Furthermore, in the hospitality industry, many studies have analyzed the impact of online customer feedback in sales. However, within the restaurant context, research is still scarce (W. G. Kim et al., 2016). When compared with state-of-the-art, the proposed approach presents several advantages. Firstly, we contribute to such gap by showing how ORs impact on restaurant performance. Secondly, the usefulness of data analytics to combine information from different sources is reflected into the perceived value of business information for restaurateurs to prioritize actions efficiently. Thus, this innovative approach provides a sales forecast with a good precision and that can be further tested in other hospitality contexts.

As for practical implications, the dashboard proposed provides an overall picture of the business that was also positively assessed by experts, which validated our proposal from both a theoretical and practical perspectives. Additionally, we have some interesting findings: we concluded that both the number of customers per day and yield influence negatively the Rating given by the customer. In fact, these findings are consistent with work of Hwang (2008), who argued that a high number of customers increases the waiting time and customers tend to be more satisfied when they have experienced a shorter waiting time prior to being served. The findings are also supported by the study of Kim and Lee (2006) who concluded that price enhances customer satisfaction. Our findings also indicate that food and service quality are important and are highly mentioned in less positive reviews, which is consistent with the studies by Ryu et al. (2012) and by Golani et al. (2017).

5.2. Limitations and future research

While experts gave positive feedback, we agree that this study presents several limitations. Findings cannot be generalized beyond the scope of the analyzed case, as the analyzed sample is rather small (including just six restaurants) and from the same geographic

market, it limits the results. Also, the method to calculate the model parameters can be improved by considering other sources of information to improve forecast performance (i.e., market data, or digital innovation factors). To address these limitations, in future work we intend to increase the number of restaurants analyzed by considering, in particular, restaurants from other geographic areas, with different seasonal variations. Furthermore, the type of restaurant is also another important factor. In the future, we aim to explore data from more restaurants, particularly covering distinct types of restaurants, which would allow us to collect feedback from more restaurateurs. Moreover, we only studied reviews written in English, thus we plan to extend this study by considering reviews written in other languages. Finally, the use of data from TripAdvisor limited our customer sample only to the digital customers of this platform.

References

- Aakash, A., & Gupta Aggarwal, A. (2020). Assessment of Hotel Performance and Guest Satisfaction through eWOM: Big Data for Better Insights. *International Journal of Hospitality & Tourism Administration*, 1–30.
- Aggarwal, C. C., & Zhai, C. (2012). Mining Text Data. In *R News*. Springer Science & Business Media. <https://doi.org/10.1007/978-1-4614-3223-4>
- Bass, F. (1969). A new product growth for model consumer durables. *Management Science*, 15(5), 215–227.
- Bass, F. M. (2004). Comments on “A New Product Growth for Model Consumer Durables The Bass Model.” *Management Science*, 50(12_supplement), 1833–1840. <https://doi.org/10.1287/mnsc.1040.0300>
- Batista, F., & Ribeiro, R. (2013). Sentiment analysis and topic classification based on binary maximum entropy classifiers. *Procesamiento Del Lenguaje Natural*, 50, 77–84.
- Berezina, K., Bilgihan, A., Cobanoglu, C., & Okumus, F. (2015). Understanding Satisfied and Dissatisfied Hotel Customers: Text Mining of Online Hotel Reviews. *Journal of Hospitality Marketing and Management*, 25(1), 1–24. <https://doi.org/10.1080/19368623.2015.983631>
- Boo, H. V. (2017). Service Environment of Restaurants: Findings from the youth customers. *Journal of ASIAN Behavioural Studies*, 2(2), 67. <https://doi.org/10.21834/jabs.v2i2.183>
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0 Step-by-step data mining guide. *CRISP-DM Consortium*.
- Chern, C. C., Wei, C. P., Shen, F. Y., & Fan, Y. N. (2015). A sales forecasting model for consumer products based on the influence of online word-of-mouth. *Information Systems and E-Business Management*, 13(3), 445–473. <https://doi.org/10.1007/s10257-014-0265-0>
- Cortez, P. (2014). *Modern optimization with R*. Springer.

- Costa, A., Guerreiro, J., Moro, S., & Henriques, R. (2019). Unfolding the characteristics of incentivized online reviews. *Journal of Retailing and Consumer Services*, 47, 272–281.
- Croes, R., & Ridderstaat, J. (2017). The effects of business cycles on tourism demand flows in small island destinations. *Tourism Economics*, 23(7), 1451–1475. <https://doi.org/10.1177/1354816617697837>
- Cross, R. G., Higbie, J. A., & Cross, D. Q. (2009). Revenue management's renaissance: A rebirth of the art and science of profitable revenue generation. *Cornell Hospitality Quarterly*, 50(1), 56–81.
- Fan, Z. P., Che, Y. J., & Chen, Z. Y. (2017). Product sales forecasting using online reviews and historical sales data: A method combining the Bass model and sentiment analysis. *Journal of Business Research*, 74, 90–100. <https://doi.org/10.1016/j.jbusres.2017.01.010>
- Feinerer, I., & Horik, K. (2018). *tm: Text mining package (Version 0.7-6)*.
- Fernández-Miguélez, S. M., Díaz-Puche, M., Campos-Soria, J. A., & Galán-Valdivieso, F. (2020). The impact of social media on restaurant corporations' financial performance. *Sustainability*, 12(4), 1–14. <https://doi.org/10.3390/su12041646>
- Few, S. (2006). *Information Dashboard Design - The Effective Visual Communication of Data*. O'Reilly.
- Filho, D., Paranhos, R., Rocha, E., Batista, M., Jr., J., Santos, M., & Marino, J. (2013). When is statistical significance not significant? *Brazilian Political Science Review*, 7(1), 31–55.
- Gan, Q., Ferns, B. H., Yu, Y., & Jin, L. (2016). A text mining and multidimensional sentiment analysis of online restaurant reviews. *Journal of Quality Assurance in Hospitality & Tourism*, 465–492.
- Golani, N., Manglik, A., & Pawar, P. S. (2017). Factors Influencing Customer Satisfaction & Customer Delight in Fine Dining Restaurants. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, 5(II), 629–638.
- Guerreiro, J., & Moro, S. (2017). Are Yelp's tips helpful in building influential consumers? *Tourism Management Perspectives*, 24. <https://doi.org/10.1016/j.tmp.2017.08.006>
- Han, H., Mankad, S., Gavirneni, N., & Verma, R. (2016). What Guests Really Think of Your Hotel : Text Analytics of Online Customer Reviews. *Cornell Hospitality Report*, 16(2), 3–17.
- Hauke, J., & Kossowski, T. (2011). Comparison of values of pearson's and spearman's correlation coefficients on the same sets of data. *Quaestiones Geographicae*, 30(2), 87–93. <https://doi.org/10.2478/v10117-011-0021-1>
- Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? *Journal of Interactive Marketing*, 18(1), 38–52.
- Heo, C. Y. (2017). New performance indicators for restaurant revenue management: ProPASH and ProPASM. *International Journal of Hospitality Management*, 61, 1–3.
- Hornik, K., & Hornik, M. K. (2018). *Package 'NLP'*. <http://cran.salud.gob.sv/web/packages/NLP/>
- Hsu, S. H., Hsiao, C. F., & Tsai, S. B. (2018). Constructing a consumption model of fine dining from the perspective of behavioral economics. *PLoS ONE*, 13(4), 1–21. <https://doi.org/10.1371/journal.pone.0194886>
- Hwang, J. (2008). Restaurant table management to reduce customer waiting times. *Journal of Foodservice Business Research*, 11(4), 334–351. <https://doi.org/10.1080/15378020802519603>
- Irfan, R., King, C. K., Grages, D., Ewen, S., Khan, S. U., Madani, S., Kolodziej, J., Wang, L., Chen, D., Rayes, A., Tziritas, N., Xu, C., Zomaya, A., Alzahrani, A., & Li, H. (2019). A survey on text mining in social networks. *The Knowledge Engineering Review*, 30(2), 157–170. <https://doi.org/10.1017/S0269888914000277>

- Jalilvand, M. R., Salimipour, S., Elyasi, M., & Mohammadi, M. (2017). Factors influencing word of mouth behaviour in the restaurant industry. *Marketing Intelligence & Planning*.
- Jovi, A., Brki, K., & Bogunovi, N. (2015). An overview of free software tools for general data mining. *37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 1112–1117. <https://doi.org/10.1109/MIPRO.2014.6859735>
- Kao, A., Poteet, S. R., & Eds. (2007). *Natural language processing and text mining*. Springer Science & Business Media.
- Kim, S., & Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 32(3), 669–679. <https://doi.org/10.1016/j.ijforecast.2015.12.003>
- Kim, W. G., & Lee, Y. (2006). Predictors of relationship quality and relationship outcomes in luxury restaurants. *Journal of Hospitality & Tourism Research*, 30(2), 143–169. <https://doi.org/10.1177/1096348005285086>
- Kim, W. G., Li, J. J., & Brymer, R. A. (2016). The impact of social media reviews on restaurant performance: The moderating role of excellence certificate. *International Journal of Hospitality Management*, 55, 41–51. <https://doi.org/10.1016/j.ijhm.2016.03.001>
- Kim, W., Lim, H., & Brymer, R. (2015). The effectiveness of managing social media on hotel performance. *International Journal of Hospitality Management*, 44, 165–171. <https://www.sciencedirect.com/science/article/pii/S0278431914001704>
- Kimes, S. (1999). Implementing restaurant revenue management: A five-step approach. *Cornell Hotel and Restaurant Administration Quarterly*, 40(3), 16–21.
- Knutson, B., Stevens, P., & Patton, M. (2008). DINESERV: Measuring Service Quality in Quick Service, Casual/Theme, and Fine Dining Restaurants. *Journal of Hospitality & Leisure Marketing*, 3(2), 35–44.
- Kwok, L., Xie, K. L., & Richards, T. (2017). Thematic framework of online review research. *International Journal of Contemporary Hospitality Management*, 29(1), 307–354.
- Lai, W.-H. (2017). Modeling the Technology Diffusion by using Bass Model. *Proceeding of AC 2017*, 169–195.
- Lasek, A., Cercone, N., & Saunders, J. (2016). Restaurant Sales and Customer Demand Forecasting: Literature Survey and Categorization of Methods. In *International Summit, Smart City 360°* (pp. 479–491). Springer, Cham.
- Maggs, G. (2020). *The evolution of Europe's financial stability agenda (Doctoral dissertation)*. University of the West of England (Bristol).
- Mahajan, V., Muller, E., & Bass, F. M. (1991). New Product Diffusion Models in Marketing: A Review and Directions for Research. *Diffusion of Technologies and Social Behaviour*, 125–177.
- Massiani, J., & Gohs, A. (2015). The choice of Bass model coefficients to forecast diffusion for innovative products: An empirical investigation for new automotive technologies Research in Transportation Economics The choice of Bass model coefficients to forecast diffusion for innovat. *Research in Transportation Economics*, 50(February 2018), 17–28. <https://doi.org/10.1016/j.retrec.2015.06.003>
- McGuire, K. A. (2016). *Hotel Pricing in a Social World*. John Wiley & Sons.
- Miner, G., Elder IV, J., & Hill, T. (2012). *Practical text mining and statistical analysis for non-structured text data applications*. Academic Press.
- Moro, S., Lopes, R. J., Esmerado, J., & Botelho, M. (2020). Service quality in airport hotel chains through the lens of online reviewers. *Journal of Retailing and Consumer Services*, 56, 102193.
- Moro, S., Ramos, P., Esmerado, J., Mohammad, S., & Jalali, J. (2019a). Can we trace back hotel online reviews' characteristics using gamification features? *International Journal of*

- Information Management*, 44, 88–95. <https://doi.org/10.1016/j.ijinfomgt.2018.09.015>
- Moro, S., Ramos, P., Esmerado, J., Mohammad, S., & Jalali, J. (2019b). Can we trace back hotel online reviews' characteristics using gamification features? *International Journal of Information Management*, 44, 88–95. <https://doi.org/10.1016/j.ijinfomgt.2018.09.015>
- Mukaka, M. M. (2012). A guide to appropriate use of Correlation coefficient in medical research. *Malawi Medical Journal*, 24(3), 69–71. <https://doi.org/10.1016/j.cmpb.2016.01.020>
- Neirotti, P., Raguseo, E., & Paolucci, E. (2016). Are customers' reviews creating value in the hospitality industry? Exploring the moderating effects of market positioning. *International Journal of Information Management*, 36(6), 1133–1143.
- Nitiwanakul, W. (2014). A Comparative Study of Customer Perceived Value as a Driver for Fine Dining Restaurant Selection: A Case of Thai Consumers and Expatriates. In *AU Journal of Management* (Vol. 12, Issue 1, pp. 1–13). <http://repository.au.edu/handle/6623004553/14670>
- O'Connor, P. (2010). Managing a Hotel's Image on TripAdvisor. *Journal of Hospitality Marketing & Management*, 19(7), 754–772.
- Pestana, M., Pereira, R., & Moro, S. (2018). A Productivity Dashboard for Hospitals: An Empirical Study. *UroSymposium on Systems Analysis and Design*, 184–199. <https://doi.org/10.1007/978-3-030-00060-8>
- Pestana, M., Pereira, R., & Moro, S. (2020). Improving Health Care Management in Hospitals Through a Productivity Dashboard. *Journal of Medical Systems*, 44(4), 1–19.
- Phillips, P., Barnes, S., Zigan, K., & Schegg, R. (2017). Understanding the Impact of Online Reviews on Hotel Performance: An Empirical Analysis. *Journal of Travel Research*, 56(2), 235–249. <https://doi.org/10.1177/0047287516636481>
- Pimpão, P., Correia, A., Duque, J., & Zorrinho, C. (2016). Diffusion Patterns in Loyalty Programs'. *Tourism and Hospitality Management (Advances in Culture, Tourism and Hospitality Research, Volume 12)*. Emerald Group Publishing Limited, 115–126.
- Quan, Q., & Cai, K. Y. (2009). Additive decomposition and its applications to internal-model-based tracking. *Proceedings of the 48th IEEE Conference on Decision and Control (CDC) Held Jointly with 2009 28th Chinese Control Conference*.
- Rogers, E. M. (2003). *Diffusion of Innovations* (5th edition). Free Press.
- Ryu, K., Lee, H., & Gon Kim, W. (2012). The influence of the quality of the physical environment, food, and service on restaurant image, customer perceived value, customer satisfaction, and behavioral intentions. *International Journal of Contemporary Hospitality Management*, 24(2), 200–223. <https://doi.org/10.1108/09596111211206141>
- Saad, M., AbuKhalifeh, A. N., Slamati, S. S., & TengkuYacob, T. N. F. F. (2020). Assessing the Use of Linear Regression Analysis in Examining Service Quality and Customer Satisfaction Relationship in Premium Casual Restaurants (PCR) in Subang Jaya (Klang Valley) Malaysia. *Review of Integrative Business and Economics Research*, 9, 369–379.
- Silge, J., & Robinson, D. (2017). *Text Mining with R - A Tidy Approach*. O'Reilly Media, Inc.
- Sparks, B. A., So, K. K. F., & Bradley, G. L. (2016). Responding to negative online reviews: The effects of hotel responses on customer inferences of trust and concern. *Tourism Management*, 53, 74–85.
- Spearman, C. (1904). The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*, 15(1), 72–101. <https://doi.org/10.1037/h0065390>
- Spyridou, A. (2017). Perceived Service Quality and Customer Revisiting Intention: The Case of "all you can eat" Asian Restaurants in Southern Taiwan. *Journal of Tourism, Heritage & Services Marketing*, 3(2), 30–38.
- Sultan, F., Farley, J., & Lehman, D. (1990). A meta-analysis of applications of diffusion models. *Journal of Marketing Research*, 27, 70–77.

- Swinscow, T., & Campbell, M. (2002). *Statistics at square one* (Tenth Edit). BMJ.
- The KPI Institute, . (2016). *Top 25 Restaurant KPIs of 2016 Extended Edition* (A. Brudan (ed.)).
- Thompson, G. M. (2010). Restaurant profitability management: The evolution of restaurant revenue management. *Cornell Hospitality Quarterly*, 51(3), 308–322.
- True, L. (2018). Package ‘*sentimentr*.’ <https://cran.r-project.org/web/packages/sentimentr/>
- UNWTO. (2018). *Tourism Highlights 2018 Edition*.
<https://doi.org/https://doi.org/10.18111/9789284419876>
- Wang, Y., & Rodgers, S. (2011). Electronic Word of Mouth and Consumer Generated Content: From Concept to Application. In *Handbook of Research on Digital Media and Advertising: User Generated Content Consumption* (pp. 212–231). IGI Global.
<https://doi.org/10.4018/978-1-60566-792-8.ch011>
- Weatherford, L. R., & Kimes, S. E. (2003). A comparison of forecasting methods for hotel revenue management. *International Journal of Forecasting*, 19(3), 401–415.
- Wexler, S., Shaffer, J., & Cotgreave, A. (2017). *The Big Book of Dashboards: Visualizing Your Data Using Real-World Business Scenarios*. John Wiley & Sons.
- 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*, 44, 120–130. <https://doi.org/10.1016/j.ijhm.2014.10.013>
- Xie, K. L., Zhang, Z., & Zhang, Z. (2014). The business value of online consumer reviews and management response to hotel performance. *International Journal of Hospitality Management*, 43, 1–12.
- Xie, K., So, K. K. F., & Wang, W. (2017). Joint effects of management responses and online reviews on hotel financial performance: A data-analytics approach. *International Journal of Hospitality Management*, 62, 101–110.
- Yaffee, R., & McGee, M. (2000). *An introduction to time series analysis and forecasting: with applications of SAS and SPSS*. Elsevier.
- Yang, W., & Mattila, A. S. (2016). Why do we buy luxury experiences? Measuring value perceptions of luxury hospitality services. *International Journal of Contemporary Hospitality Management*, 28(9), 1848–1867.
- Yu, X., Liu, Y., Huang, X., & An, A. (2012). Mining online reviews for predicting sales performance: A case study in the movie domain. *IEEE Transactions on Knowledge and Data Engineering*, 24(4), 720–734.
- Zhang, J., & Verma, R. (2017). What Matters Most to Your Guests: An Exploratory Study of Online Reviews. *Cornell Hospitality Report*, 17(4), 3–13.
<http://scholarship.sha.cornell.edu/chrpubs%0Ahttps://sha.cornell.edu/faculty-research/centers-institutes/chr/research-publications/documents/rohit-verma-online-reviews.pdf>
- Zhang, Tian, Y.-X., & Fan, L.-W. (2020). Improving the Bass model’ s predictive power through online reviews, search traffic and macroeconomic data. *Annals of Operations Research*, 1–42. <https://doi.org/10.1007/s10479-020-03716-3>