LUISS

Department
of Business and Management

Ph.D in Ph.D in Management

Cycle Cycle 34

Essays on Applications of Sentiment Analysis and Text Mining Techniques in Online Reviews

Feray Adiguzel

SUPERVISOR

Carmela Donato

CO-SUPERVISOR

Moamen Elsherbiny

CANDIDATE

Academic Year 2022/2023

Essays on Applications of Sentiment Analysis and Text Mining Techniques in Online Reviews

By

Moamen ELSHERBINY

A thesis submitted in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY IN MANAGEMENT

LUISS University

2023

DEDICATION

I dedicate this dissertation to my parents and my wife. Without their tremendous encouragement in the past few years, it would be impossible for me to complete my study. Thank you for everything you did for me!

List of Abbreviations

Abbreviations	Full term
WOM	Word-Of-Mouth
LDA	Latent Dirichlet Allocation
OCR	Online Consumer Reviews
OPR	Online Product Reviews
NLP	Natural Language processing
eWOM	Electronic Word-Of-Mouth
VADER	Valence Aware Dictionary and sEntiment Reasoner
BoW	Bag of Word
NLTK	Natural Language Toolkit
OR	Online Ratings
NRC Emotion Lexicon	Word-Emotion Association
ECT	The expectancy-confirmation theory

ACKNOWLEDGEMENTS

Completing my Ph.D. has been a wonderful and challenging journey. I am deeply grateful to many people for their kind assistance during this time. First, I would like to express my appreciation to my PhD supervisors; Feray Adiguzel and Carmela Donato for all their valuable support and guidance. Without their fruitful feedback, this dissertation would not have been possible.

Feray Adiguzel has played an important role in my Ph.D. journey. I feel blessed to have the opportunity to learn a lot from her. I am thankful to her for her guidance, flexibility, and her support in completing my thesis. Carmela Donato has been a great support, her insights helped me write this work.

I extend my sincere gratitude to my dissertation committee members, Prof Csilla Horvath and Prof Gomaa AGAG, for their constructive feedback and helps in improving the quality of the dissertation. I would also like to thank my friends and colleagues. Finally, I would like to thank my family and my wife who have always supported me during this period. They did everything possible to see me here and today I want to say thank you for everything.

Author's Declaration

I hereby certify that the information in this thesis has not previously been submitted for a degree at this university or any other. I further declare that only my own research fulfilled the framework for this thesis.

Moamen ELSHERBINY

Publications

Journal Articles

- Social Influence in Hotel Reviews: The Effect of Emotions and Hotel Characteristics (under the review process - *Journal of Service Research*)
- META-ANÁLISIS SOBRE SENTIMENT ANALYSIS Y PUNTUACIONES DIGITALES DE LOS CLIENTES EN EL SECTOR HOTELERO" VI FORO Internacional de Turismo Maspalomas CostaCanaria (Gran Canaria, Spain) 2018 [ISBN : 978-84-9042-342-4]
- "Customer Sentiment and Online Customer Ratings in Hospitality: A Gap in Tourism Studies" International Tourism Studies Association (ITSA) and International Tourism Educators South Africa (TESA) (Tshwane, South Africa) 2018 [pp. 139-143]

Conference Papers

- "Emotions in Review Text and Social Influence", 4th ISMS Marketing Science Conference 2021
- "Topic Modelling Application for Luxury Hotel Reviews", 25th MMRA Marketing Congress, Ankara, Turkey 2021. Published in proceeding.Marketing Congress 2021
- "The Relationship between Customer Sentiment and Online Customer Ratings for Chain Hotels in Gran Canaria" Spring Symposium on Challenges in Sustainable Tourism Development (SSTD) (Gran Canaria, Spain) 2018 [ISBN: 978-84-09-03853-4]

- “Electronic Word-of-Mouth as a tool to Destination Management Development” Spring Symposium on Challenges in Sustainable Tourism Development (SSTD) (Gran Canaria, Spain) 2017 [ISBN: 978-84- 697-7603-2]

PhD Visiting Research

- PhD visiting researcher at HEC Paris – Marketing Department (from September 2021 – end of January 2022), Paris-France

Teaching Assistant

- TA in Hotel Management at LUISS University – Dr. Youngah Guahk (2020)
- TA in Tourism Management at LUISS University – Dr. Youngah Guahk (2020)
- TA in Corporate Strategy at LUISS University – Dr. Evangelos Syrigos (2021)

Abstract

Online reviews consider the most important issue that concerns hospitality managers due to their influence on customers. Online reviews have a significant impact on organizations' profitability as well as the consumers' decision making. Therefore, understanding the factors that influence consumer online posting behavior is curial for managers. This dissertation consists of three papers that mainly focuses on the context of hotels. The purpose of this dissertation is to investigate how social influence (i.e., prior average rating) affects subsequent ratings and to what extent expressed emotions by reviewers (i.e., joy and anger) and hotel characteristics (i.e., price category and affiliation) can affect this relationship. In addition, this dissertation is focused on the most important topics that consumers are talking about by investigating the impact of topics and emotions embedded in reviews on customer satisfaction.

Using an econometric and text mining method based on online review data from TripAdvisor, Study 1 explores the relationship between social influence (i.e., the prior average rating) and subsequent ratings, as well as the impact of expressed emotions by reviewers (such as joy and anger) and hotel characteristics (such as price category and affiliation).

The results of this study indicate that when a review express is joy, the positive relationship between social influence and subsequent ratings becomes less positive; contrarily, when a review express is anger, the relationship becomes more positive.

Moreover, the negative effect of budget (vs mid-priced) hotels on subsequent ratings is more negative with increasing prior average ratings and the effects of luxury (vs mid-priced) and independent (vs chain) hotels on subsequent ratings do not change with prior average ratings, indicating no social influence effect.

Using sentiment analysis and topic modeling methods based on online review data, Study 2 first extracts topics in review texts and examines the influence of topics and other review-specific variables such as emotions embedded in reviews on satisfaction ratings for luxury chain hotel reviews. The findings of this study reveal that review topics with pay problems and limited hotel amenities have low satisfaction ratings. Additionally, review polarity, joy, prior average rating, and distribution of average rating are positively associated with online ratings whereas review length is negatively associated. Finally, using online review data from TripAdvisor, Study 3 explores the effect of the prior mode rating—or the rating score that has the highest frequency of satisfaction ratings—on individual ratings. The findings of this study imply that the positive effect of the prior mode ratings and subsequent ratings is moderated by reviewer's geographical proximity and reviewer's expertise where the relation is stronger for reviewers located close to the hotel and weaker for reviewers who have higher level of contributions.

This dissertation contributes to the hospitality marketing literature by providing new theoretical insights. Moreover, the empirical results of this dissertation also reveal significant managerial implications for hospitality businesses and online review communities.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

GENERAL INTRODUCTION

1.1 Research Background

Online reviews (OR) have grown in popularity as an important source of word-of-mouth (WOM) that can influence product sales and profitability due to the emergence of the internet and social media (Chevalier et al., 2006; Ye et al., 2009; Zhu et al., 2010). Online reviews are the voice of the customer, they provide important insights to managers and offer a chance to hear from customers. Online reviews are more accessible and cost-free as a research tool than in-depth interviews or focus groups (Murgado et al., 2012). Understanding the factors that influence consumers' online review-posting behavior is therefore vital for business success and theoretical advancement. Although there is a growing body of academic studies on this topic, little is known about effect of social influence on reviewers' satisfaction ratings.

The decision-making process of consumers is significantly influenced by online customer reviews and ratings, which are regarded as a very important source of information (Chevalier et al., 2006). Online product reviews (OPR) are frequently read by customers before making purchases given that experience goods' quality can only be judged after consumption. Online reviews (OR) play an even more significant role in the marketing and the hospitality sector as a decision-making tool. However, online reviews can provide consumers with unbiased information (Muchnik et al., 2013; Zhang et al., 2016; and Wang et al., 2018).

According to marketing charts in 2007, around 82% of consumers claimed that online product reviews directly influenced their purchasing decisions, and over 75% of them believed that recommendations from people who had actually used the product themselves were the most reliable information sources (Nielson 2007).

After the purchasing process, they might also be exposed to personal testimonies written by previous consumers on a product review page. Academics are commonly believing that a person's online review behavior won't be influenced by prior reviews until a product has been purchased and consumed. Moe et al., (2012) and Schlosser (2005) both indicated that when consumers are deciding to rate products, they frequently look to the ratings of previous customers and adjust their own ratings accordingly. However, the impact of previous reviews may also be applied when someone reads online reviews prior to making a purchase, which can affect the pre-purchase expectations of a product.

Therefore, the purpose of my dissertation is to use text mining and machine learning techniques to analyze the social influence effect in online reviews to determine:

- The effect of prior average ratings as social influence on subsequent ratings and the moderating roles of both emotions embedded in text and hotel characteristics (Paper 1).
- Product/service failures and brand positioning problems for instance, luxury chain hotels, using topic modelling and sentiment analysis, (Paper 2).
- The effect of the mode of prior ratings on rating behaviors and the moderating roles of reviewer's expertise and geographical proximity (Paper 3).

1.2 Research Significance

Previous research has claimed that online reviews are reliable and unbiased (Hu, Liu, & Sambamurthy, 2011). However, a growing body of literature disputes that review rating environments, such as prior average review ratings and variance in prior ratings, have an impact on consumers' online review behaviour (Ho et al., 2017; Lee et al., 2015; Li et al., 2008; and Moe et al., 2012). This suggest that online reviews behaviors might be influenced by social factors.

For instance, consumer product experiences as well as other people's perceptions of a product may have an impact on consumer online review behavior, including the probability that they will post reviews online and how they ultimately rate the reviews. Based on an extensive review of the literature, there are several research gaps on this topic. First, previous research has provided only a limited understanding of the social influences on consumers' online review behavior, particularly regarding the different factors that may strengthen or weaken this phenomenon. Second, even though previous research has shown that consumer product experiences and the prior average ratings may affect how consumers behave when posting reviews online, moderating effects in hotel reviews have hardly ever been discussed.

1.3 Research Framework

Three main studies are included in this dissertation. These studies are focus on hotel products rather than manufactured ones as they are more experience-oriented and have intangible, variable, and perishable qualities. As a result, it is more likely that online hotel reviews will be influenced by social factors.

Using text mining technique based on online review data from TripAdvisor, Study 1 investigates the relationship between social influence (i.e., the prior average rating) and subsequent ratings, as well as the impact of expressed emotions by reviewers (such as joy and anger) and hotel characteristics (such as price category and affiliation). This study investigated the following research questions: *(1) What is the role of emotions embedded in review texts on subsequent hotel review ratings? (2) How do product characteristics affect the individual rating behavior?*

Using topic modeling and sentiment analysis, Study 2 extracts topics in review texts and examines the influence of topics and other review-specific variables such as emotions embedded in reviews on satisfaction ratings for luxury chain hotel reviews. This study examined the following research questions*: (1) What topics emerge in online reviews from guests of luxury hotels? (2) How do these issues, along with other review elements like the sentiments expressed in the review text, impact satisfaction scores? (3) How do these topics vary across three luxury hotels?*

In addition, Study 3 explores the phenomena of social influence effect at post-purchase stage on advancing our understanding of how consumers influenced by different aspects of rating characteristics such as the mode of prior ratings — or the rating score that has received the most satisfaction votes. This study investigated the following research questions: *(1) What is effect of mode of prior ratings on post-purchase consumer satisfaction assessments (i.e., subsequent ratings)? (2) What is moderating role of the reviewer's geographical proximity on the relationship between prior mode rating and subsequent ratings? (3) What is the moderating role of the reviewer's expertise on the relationship between prior mode rating and subsequent ratings?*

This dissertation is supported by a set of fundamental theories, including:

1. *Social influence theory:* people can confirm or deviate from others depending on their social needs (conformity or uniqueness) (Fromkin, 1970, Sherif, 1936) or normative conflict conditional on situational characteristics (Sridhar & Raji Srinivasan, 2012). In social influence theory (Fromkin, 1970), group consensus is used by people to form their own opinions; thus, a main positive effect of prior average rating is also expected in hotel reviews context. The focus of this dissertation is on how consumers' online review behavior is influenced by prior average review ratings, therefore social influence is used as a core theory.

2. *Mobilization-minimization theory (Taylor, 1991):* consumers respond to negative emotions with an initial phase of rejection (i.e., mobilization) of the stimulus that generated the negative emotion followed by a subsequent phase of cognitive effort aimed at reducing the undesired effects of the negative emotion (i.e., minimization). Thus, after a joyful consumer experience, when consumers see high prior average ratings, they reduce their ratings, which might cause a decreasing trend over time for a hotel. However, reviews with anger emotions might adjust and increase their subsequent ratings due to a reduction in their negative feelings and in order to feel better. Therefore, mobilization-minimization theory is employed accordingly.

3. *The Expectancy-Confirmation Theory (Oliver, 1980):* ECT has been used for understanding customer satisfaction. According to this theory, if customers' perceived quality is higher than their expectation, customers are satisfied, and, conversely, if a service experience does not match or exceed customers' expectations, they will feel dissatisfied. For hotel customers, one way to show their satisfaction/dissatisfaction is through their evaluations of products or services through online customer ratings.

CHAPTER 2

Social Influence in Hotel Reviews: The Effect of Emotions and Hotel Characteristics

2.1 Introduction

According to prior literature (e.g., Zhu & Zhang, 2010; Yacouel & Fleischer, 2012), high volume of positive reviews increases sales and allows hotels to ask for a premium price. Online reviews are especially eminent in the hospitality and tourism sectors. Indeed, Vermeulen and Seegers (2009) indicated that the impact of online reviews in online travel purchases per year is over $10 billion. Even though online review ratings are supposed to be an unbiased source of information for consumers, a rising number of literatures has demonstrated the presence of a so-called "social influence effect," since being exposed to others' average review ratings influences consumers' online rating behavior (Schlosser, 2005; Li & Hitt, 2008; Moe & Schweidel, 2012; Lee, Hosanagar & Tan, 2015; Ho, Wu & Tan, 2017). In this paper, social influence is defined as reviewers' tendency to alter or make adjustments to their rating choices (**either intentionally or unintentionally)** due to exposure to prior average ratings (Sridnar & Srinavasan, 2012). Consequently, when making a rating decision, reviewers are anchoring a prior average rating and make changes in their own rating depending on perceived (dis)confirmation from their own consumption experience (Adomavicius, Bockstedt, Curley, & Zhang, 2013; Li, Zhang, Meng & Zhang, 2019), thereby employing an anchoring-and-adjustment heuristic (Tversky & Kahneman, 1974).

The existing literature is mostly focused on detecting the presence of and finding the reasons for social influence, emphasizing reviewer characteristics as a moderator for product reviews (Schlosser, 2005; Hu & Li, 2011; Moe & Schweidel, 2012; Lee, Hosanagar & Tan, 2015; Zhang, Zhang, & Yang, 2016), in addition to restaurant reviews (Ma, Khansa, Yun, & Sung, 2014; Guo & Zhou, 201; Li, Zhang, Meng, & Zhang, 2019; Li, Meng, Jeong, & Zhang, 2020). However, apart from a few studies in the marketing literature (e.g., Sridhar & Srinivasan, 2012), the moderators of social influence in hotel reviews have largely been neglected, especially in the prior hospitality literature. According to Brehm (1966), individuals' likelihood of conformity or non-conformity behavior depends on emotions and dispositions. As regards emotions, since hotels are related to experience and leisure, they play a fundamental role in consumer evaluations of their hotel stays. In fact, service-encounter performance influences consumer emotions (Jani & Han, 2011) and hotel experience emotions have been shown to significantly influence consumer satisfaction (Bravo, Martinez, & Pina, 2019; Lee, Lee & Koh, 2019; Bastiaansen et al., 2019). Therefore, in this research, we propose that the emotions previously felt by consumers during their stay and expressed in a review text can be a possible moderator of the social influence effect on consumer rating behaviors.

Even though emotional responses in hedonic service situations —e.g., tourism, entertainment, and luxury, have been emphasized in the service marketing literature (Arnould & Price, 1993; Bigné, Mattila, & Andreu, 2008; Jani & Han, 2015), their impact on individual rating behavior has not yet been investigated. In this study, joy, and anger, being the most frequently investigated emotions in online reviews, were selected from Plutchik's model of emotions (1980). Anger and Joy are considered as high-arousal emotions that have a significant influence on purchasing decisions (Achar et al., 2016).

Joy is identified as being important for perceived helpfulness and may raise favourable evaluations of products and services and promote brand loyalty (Felbermayr & Nanopoulos, 2016) whereas anger, as the opposite of joy, has been found to raise customer frustration and complaint behavior and it is a red flag for hotel managers and the most frequently expressed emotion in negative reviews (Wetzer, Zeelenberg, & Pieters, 2007). Given this fact, it's essential for companies targeting to enhance customer experiences to understand how these emotions are triggered and how they affect consumer behavior (Puccinelli et al., 2009).

Regarding dispositions (Brehm, 1966), consumers may have already formed a pre-decisional disposition toward the hotel based on quality indicators such as price, brand, and information present in previous reviews. Thus, the effect of social influence might change depending on service-experience situational factors, especially noticeable quality indicators that consumers prominently manifest during pre-booking and post-service evaluations. Consequently, we propose that the emotions previously felt by consumers during their stay and expressed in the text of a review can be a possible moderator of the social influence effect on consumers' rating behavior. Moreover, given that individual ratings differ depending on hotel characteristics (Xu, 2019; Geetha, Singha & Sinha, 2017; Banerjee & Chua, 2016; Martin-Fuentes, 2016), the social influence effect might also change with them. The most observable attributes of a hotel both when booking and when writing a review are affiliation and hotel category, which are also commonly accepted quality indicators that consumers use to form their expectations.

While chain hotels belong to a series or of a group of hotels operated by the same company or owner, independent hotels are not affiliated with any brand and are self-proprietary (Kirk, 1995; Namasivayam, Enz, & Siguaw, 2000).

Chain hotels also get benefits of confidence and loyalty through their brand. Travelers' rating preferences were found to be different for independent hotels compared to chain hotels (Banerjee & Chua, 2016). Similarly, emotions and ratings were discussed as being different in premium and budget hotel reviews (Geetha, Singha & Sinha, 2017). Consequently, hotel affiliation and price category were selected as additional moderators of the relationship between social influence and consumer rating behavior. To investigate social influence effect and moderators, we used a sample of 65215 online reviews from TripAdvisor and validated the findings with 67534 reviews from Expedia for 169 hotels in New York City. We used lexicon-based sentiment tools to extract emotions embedded in reviews and other features from a review text, and then fitted a multilevel ordinal logit model containing hotel, reviewer, and emotions, along with other review-specific characteristics. To the best of our knowledge, this is the first study to jointly model emotions embedded in reviews with hotel and reviewer characteristics. Furthermore, this study contributes to a better insight on how emotions and hotel characteristics may change the effect of social influence on subsequent rating behavior in online reviews.

Understanding the social influence effect in hotel reviews is crucial for rating platforms, hotels, and travel agency managers. Consumers might be misled by rating biases, resulting in suboptimal product decisions and false product ratings. Biased ratings have the potential to affect not just which items are purchased, but also to cause over- or under-quality signals and incorrect popularity rankings, and even to harm a company's reputation. Accordingly, rating biases caused by social influence might also decrease the value and usage reputation of platforms and booking websites.

Moreover, some hotel types might suffer greater biases than others. Therefore, understanding these biases in customer reviews is an essential and very practical research challenge.

2.2 Conceptual Model and Hypotheses Development

2.2.1 Social influence and reviewing behavior

There has been a discussion in the literature concerning the relationship between the average rating of prior reviews and subsequent ratings. According to Ma et al. (2014), the connection is positive. They contend that the average rating of prior reviews can act as a signal for future consumers to set initial expectations, which will favorably affect their post-consumption evaluations. However, without offering any justifications, Wang et al. (2015) discovered a negative relationship. The inverse link could be explained by the declining trend in product ratings (Li and Hitt, 2008). Regardless of the conflicting findings in the research, there are several reasons to prove that there is a positive relation between prior average ratings and subsequent ratings (Guo et al., 2016). According to Hu and Li (2011), when users check product online, the product's prior average rating is served as a good indicator of how effective this product to users and can be used to predict the product quality. This allows them to create their own judgements and to determine the usefulness of the product (Hong and Pavlou, 2014). Moreover, it's possible that future reviewers will be socially influenced by prior reviewers to provide ratings that are similar to the average rating (Lee et al., 2015).

Customers are exposed to different social cues, and prior average ratings, dispersion in ratings (i.e., variance in ratings), and the number of prior reviews (i.e., volume) (Li et al., 2020; Li et al., 2019; Gavilan, Avello, Martinez-Navarro 2018; Guo & Zhou, 2016) can affect their subsequent rating behavior.

Despite the fact some studies have found a negative effect deriving from prior rating exposure (Moe and Trusov, 2011; Hu & Li, 2011), the literature seems to concur on the positive relationship between average prior ratings and subsequent ratings (Sridhar & Raji Srinivasan, 2012; Ma et al., 2014; Guo & Zhou, 2016; Gao et al. 2018; Li et al., 2020). This paper therefore investigates these conflicting results in the hotel review context. In social influence theory (Fromkin, 1970), group consensus is used by people to form their own opinions; thus, a main positive effect of prior average rating is also expected in hotel reviews context.

According to social influence theory, people can confirm or deviate from others depending on their social needs (conformity or uniqueness) or normative conflict conditional on situational characteristics (Sridhar & Raji Srinivasan, 2012). Conformity needs involve the tendency to adapt to others' opinions, since following others can result in fewer mistakes, which reduces both mental effort and the fear of losing one's reputation when deviating from the majority (Cialdini, 2009). Uniqueness needs relate to an individual's pursuit of differentness relative to others (Fromkin, 1970); therefore, people who have a great need for uniqueness are more likely to deviate from others (Snyder & Fromkin, 1980).

Normative conflict regards a large discrepancy between group norms and consumer experiences (Hornsey, Oppes & Svensson, 2002), which might cause deviation from social group consensus (Hornsey et al., 2002).

Transposing these findings on rating behavior, people either tend to rate higher when they see high prior ratings or, conversely, they can go in the opposite direction from others' average ratings; this therefore implies either a positive or a negative effect of social influence on rating behavior (Adomavicius et al., 2013). Additionally, people may reduce their rating if they feel the need to be discriminatory (Schlosser, 2005), but they may also raise it if others' ratings are very low (Moe & Trusov, 2011).

Studies on social influence through prior average ratings are summarized in Table 2.1 The research by Schlosser (2005) on the causes of social influence bias proved that exposure to others' ratings caused a decrease in individual rating due to triggered social concerns about self-presentation (appearing more knowledgeable and competent) or the desire to avoid being regarded as discriminatory. While the work of both Moe and Trusov (2011) and Muchnik, Aral and Taylor (2013) illustrated that disclosed prior ratings triggered significant bias in rating behavior, Muchnik et al. (2013) also found that the biased direction changed with the sentiment of the prior rating. In other words, people prefer to follow positive opinions and are skeptical of negative ones in the content of social news aggregation websites. In addition, Muchnik et al., (2013) concluded that some categories or products (in their data, news about "politics, culture and society and business") are more vulnerable to social influence bias than others. According to Sridhar and Raji Srinivasan (2012), other consumers' online ratings reduce the effects of positive and negative elements of product experience.

A few studies that examined moderators of the impact of social influence on subsequent ratings found that social influence is higher for consumers who invest in less cognitive effort, for those with moderate product experience, for non-elite reviewers, and when the temporal distance between product experience and review posting is large (Li, Meng, Jeong & Zhang, 2020; Li, Zhang, Meng & Zhang, 2019). Finally, Xue, Dong, Gao, Yu, and Taras (2020) illustrated the finding that the herding behavior of international reviewers, in terms of rating deviation from prior average ratings, increases when cultural and geographical distance increases, and their effects are weak for experienced travellers.

Table 2.1. Overview of literature and contributions

Source, Year, context, method, model	IV and DV	Moderator(M)
This study **Hotel reviews on TripAdvisor& Expedia,** **Multilevel ordinal logit**	**IV: Prior average rating** **DV: Subsequent ratings**	**M1: Review specific: (subsequent) emotions (anger and joy), variance in prior ratings** **M2: Hotel specific: hotel price scale (budget vs luxury), ownership structure (chain vs independent)** **M3: Reviewer specific: Whether review mentions previous reviews or reviewers**
Filieri, R., Raguseo, E., & Vitari, C. (2021) Extreme Negative Rating and Review	IV: Extreme negative rating, DV: Review helpfulness	M1: Review volume M2: Average Rating Score M3: Hotel Category M4: Certificate of excellence

Helpfulness: The Moderating Role of Product Quality Signals		M5: Chain
Xue, Dong, Gao, Yu & Taras 2020, Hotel reviews on TripAdvisor, Multiple regression	IV1: Cultural distance IV2: geographic distance DV: Rating deviation (i.e., herding) Note. Average rating & prior review volume were used as control variables	M1: Hospitality (i.e., travel) experience M2: Geographical distance
Li, Meng, Jeong, & Zhang 2020, Restaurant reviews on Yelp, Ordered logit	IV: Prior average rating DV: Subsequent rating	M1: Moderate dining experience, cognitive effort, variance in prior ratings M2: Elite vs non-elite reviewers
Li, Zhang, Meng, & Zhang 2019 Restaurant reviews on Yelp, Ordered Logit	IV1: Prior average rating IV2: Volume (number of prior reviews) DV: Subsequent rating	M1: Review temporal distance (i.e., the duration between consumption and review posting)
Wang, Zhang & Hann 2018, Chinese website for books, movies, music	IV: Prior average friend's rating DV: Subsequent rating	M1: Network size M2: Age of product

Guo & Zhou 2016, Restaurant reviews on Yelp, Ordered logit	IV: Prior average rating DV: Subsequent rating	M1: Volume and variance of prior ratings M2: Subsequent reviewer's connectedness and expertise
Lee, Hosanagar &Tan 2015, Movie reviews, Rating incidence: multilevel mixed-effects probit regression Rating: multilevel mixed effects ordered probit	IV1: Prior average rating IV2: Number of prior reviews (volume) DV2: Subsequent rating DV1: Decision to rate (latent)	M1: Strangers ("crowd") versus friends, M2: Popularity M3: Audience size
Ma et al 2014, Restaurant reviews on Yelp, Field Experiment, Hierarchical/multi-level model (HLM)	IV: Prior average rating DV: Subsequent rating	M1: Reviewer's prior experience, geographical mobility, social connectedness, and gender M2: Length of review, the time interval between reviews (frequent vs non-frequent))
Munchik, Aral & Taylor 2013, social news aggregation website,	IV: Valence category (negative/positive and neutral posts)	M1: Review topic M2: Source of prior ratings (friends vs crowds)

Hierarchical bayes	DV: Up and down voting behaviour for a post	
Moe & Schweidel 2012, Bath, fragrance, and home product reviews on BazaarVoice, Hierarchical/multi-level model (HLM), Ordered probit (incidence+evaluation)	IV1: Degree of consensus or dissention in the ratings environment IV2: Overall positivity of posted product ratings, accounting for the number of ratings that have contributed to this positivity (loadings of valence, volume, variance) DV1: Decision to rate DV2: Subsequent rating	M1: Frequent vs less frequent posters (i.e., reviewers)
Sridhar& Srinivasan 2012, Hotel reviews on online travel site, Nested ordered logit (HLM i.e., multilevel)	IV: Product experience (positive and negative features, product failure and recovery) DV: Subsequent rating	M1: Prior average rating
Hu & Li 2011, Book reviews on Amazon (i.e., product), Ordered logit	IV: Prior average rating DV: Subsequent rating	M1: Popularity of the product M2: Variance of previous reviews M3: Whether a review explicitly refers to previous reviews M4: Age of the product and the reviews

Schlosser 2005, Short film reviews, Experiment, ANOVA	Study 1: IV: public (posters) vs private (lurkers) DV: Favorability rating	M1: (+, - or neutral) Review text (valence category)

We propose that reviewer ratings can also be affected by the reviewer's emotions. Consequently, the objective of the present research is to verify the effect of social influence on rating behaviour by considering the emotions expressed in a review text after service experience.

2.2.2 Emotions embedded in review texts.

Emotions, defined as a "mental state of readiness arising from cognitive appraisals of events or beliefs" (Bagozzi, Gopinath, & Nyer's 1999, p. 184), have extensively proven to influence judgment and decisions, such as customer satisfaction (Westbrook & Olivier, 1991), consumer behavior (Watson & Spence, 2007), information searches and processing (van Kleef, 2014), content of thought (Lerner & Keltner, 2000, 2001), virality (Berger & Milkman, 2012; Yan, Zhou, & Wu, 2018), and perceived helpfulness of a review (Martin, Sintsova & Pu, 2014; Yin, Bond & Zhang, 2014; Ullah, Zeb & Kim, 2015). Related to the latter point, different emotions have been studied. For example, anger and sadness have been found to have a negative effect on perceived review helpfulness, whereas fear had a positive effect (Ren & Hong, 2019).

Plutchik’s model of eight emotions (namely anger, disgust, fear, sadness, joy, trust, anticipation, and surprise) was used by Felbermayr and Nanopoulos (2016) and Wang, Tang and Kim (2019) to explain perceived review helpfulness. Trust, joy, and anticipation were identified as the most prominent emotions influencing perceived review helpfulness for product reviews (Felbermayr & Nanopoulos, 2016).

Wang, Tang and Kim (2019) found that anger, disgust, fear, sadness, joy, and trust explain the perceived helpfulness perceptions of customers for restaurant reviews. Anger-embedded review content was found to be less significantly associated with helpfulness ratings than anxiety-embedded content (Yin, Bond & Zhang, 2014). However, the former was also perceived as more persuasive (Yin, Bond, & Zhang, 2021). Moreover, while anger decreases the informational value of reviews and product evaluations, happiness was found to not influence product evaluations (Kim & Gupta, 2012, Lelieveld & Hendricks, 2021).

Finally, Chen and Farn (2020) found that readers perceived a reviewer as exerting more cognitive effort when the emotions conveyed by the review were negative (i.e., anger and fear) than positive (i.e., pride and surprise). Generally, the literature treats each emotion independently, with an emphasis on the valence and arousal of the emotion (Keltner, Ellsworth, & Edwards, 1993; Lerner & Keltner, 2000). In terms of valence, emotions can be distinguished into positive and negatives ones (Russel, 1980). In this sense, arousal represents the intensity of emotions in terms of stimulation and relaxation, providing a framework for any delineation and combination of components (Russell & Mehrabian, 1974). Consumers are affected differently by emotions of the same valence on various levels, ranging from physiological reactions to perception, taking decisions, and coping behaviors (e.g., Keltner and Horberg 2015; Yen and Chuang 2008).

This is because emotions of similar intensity can differ significantly in terms of their fundamental appraisals. As a result, numerous psychologists (e.g., Lench et al. 2011; Lerner and Keltner 2000; Tiedens and Linton 2001; Zeelenberg et al. 2008) have highlighted the significance of investigating specific emotions. Distinct emotions can also affect a consumer's behavior and judgment in very diverse ways. or more specifically, they can influence how to analyze information either heuristically or systematically (Lerner and Tiedens 2006; Tiedens and Linton 2001). Therefore, both the valence of an emotion (positive-negative) and the intensity level of its arousal (high-low) are significant components of emotional reality (Watson & Tellegen, 1985). Another important element of emotions is related to their appraisal patterns (Lerner & Keltner, 2000). Appraisals refer to the evaluative frameworks that people use to judge or make sense of events and situations (Yap, & Tong, 2009).

Prior research underscored the role of a particular emotion appraisal in affecting a decision maker's confidence in shifting from a reference point. This is *certainty appraisal* (Inbar & Gilovich, 2011), which refers to the extent to which a specific emotion provides the confidence required to cope with the uncertainty instigated by external stimuli. The relationship between emotions and certainty appraisals was investigated by Tiedens and Linton (2001), their findings showed that emotions with certainty appraisals such as (anger & joy) increased the feeling of perceived certainty in subsequent situations than emotions with uncertainty appraisals (such as fear & sadness). To put it another way, the emotions with certainty (e.g., anger & joy) is linked to heuristic processing, while the uncertainty emotions encouraged systematic thinking (Tiedens & Linton, 2001).

These findings were consistent with previous studies such as Bodenhausen et al. (1994), who also found evidence of the influence of emotions on cognitive processing. Certainty means the degree to which a person feels confident about the outcomes of an event (Kranzbühler et al., 2020). For instance, anger is connected to a high degree of certainty; it is triggered when a person feels certain about the negative effects of an incident (e.g., Lerner and Keltner 2000; Smith and Ellsworth 1985). In our research, as in that of Kim and Gupta (2012), we decided to analyze two opposite emotions based on valence-arousal dimensions: *joy* and *anger*.

Joy is a positive high-arousal emotion that reveals how people are satisfied and pleased with a product/service (Laros &Steenkamp, 2005), whereas anger is a negative high-arousal emotion defined as a feeling of displeasure and hostility that arises after being exposed to any form of injustice or attack (Bodenhausen, Sheppard, & Kramer, 1994).

Favourable consumption experiences lead to joy-embedded reviews and high ratings, whereas unfavourable consumption experiences result in anger-embedded reviews and low ratings. Both joy and anger are characterized by high certainty appraisal. Certainty in joy and anger stems from people's ability to know exactly the source of their pleasant or unpleasant feelings, respectively (e.g., a pleasurable/unpleasant experience producing joy/anger), and from people's ability to predict what they should do in order to keep/cope with their pleasant/unpleasant feeling (e.g., laud/blame). The idea that emotions produce judgments that are congruent with their underlying certainty appraisals seems to be supported by different instances of research. For example, Baas, Dreu and Nijstad (2012) demonstrated that emotions associated with certainty lead to less structured ideation than emotions that are associated with appraisal of uncertainty. On the other hand, Tiedens and Linton (2001) found that emotions associated with uncertainty result in lower stereotyping and more attention to argument quality.

Based on this information processing line of research, it can be argued that, by way of the level of certainty, both joy and anger embedded in a review text can influence the effect that social influence has on consumers' rating behavior.

Exposure to prior average ratings serves as an anchor for the reviewer's rating scale choice (Adomavicius et al., 2013). Individuals frequently estimate unknown values by starting off a prominent value, known as anchor and subsequently adjusts that information until a satisfying value is reached (Tversky & Kahneman, 1974). Then a reviewer can start with an average prior rating anchor and apply an anchoring-and-adjustment heuristic to a rating choice based on the (dis)confirmation of their expectations, which are expressed as an emotion in a review text. Emotional experiences have been proven to have a major impact on anchoring bias (Bodenhausen, Gabriel & Lineberger, 2000; Englich & Soder, 2009; Epley & Gilovich, 2006).

As regards joy, when making judgments, people in a positive mood tend to process information less comprehensively and depend on heuristics (Forgas, 1995; Schwarz, 1998). Reliance on stereotypes (Bless, Schwarz, & Kemmelmeier, 1996; Bodenhausen, Kramer, & Susser, 1994), the source credibility heuristic (e.g., Mackie & Worth, 1989), the availability heuristic (Isen & Means, 1983), scripts (Bless, Clore, et al., 1996) and other persuasion heuristics increase with happy moods. Moreover, according to Mackie and Worth (1989) joyful people appear to depend more on source cues than on systematic message quality evaluation in persuasion events.

Positive affect, according to Estrada, Isen and Young (1997), enables systematic processing of relevant or interesting data, resulting in more comprehensive and efficient problem resolution, reducing the magnitude of anchoring effects in physicians' diagnoses, which is consistent with anchoring and adjustment theories.

In line with the above-mentioned literature, we expect the emotion of joy will make the relationship between social influence and subsequent ratings weaker. Consequently, when people write reviews embedded with joyful sentiments, they might adjust less (i.e., deviate) from the anchor value of "prior average rating", thereby reducing the effect of average prior rating. This means that reviewers' evaluations are less likely to be affected by anchoring after writing joy-embedded reviews. In formal terms:

H1: The positive relationship between prior average rating and subsequent review ratings is weaker (less positive) when joy increases.

According to previous theories, angry people also seem to not process information analytically (Forgas, 1995; Lerner, Goldberg, & Tetlock, 1998; Ric, 2004; Russell, 2003). In this sense, Bodenhausen et al. (1994) demonstrated that angry information processors are influenced by heuristic cues. When explaining these results, the authors accredited this anger-induced lack of analytic processing to "*reduced motivation for thoughtful analysis of judgment-relevant information, reduced capacity for such analysis, or something else*" (Bodenhausen et al. 1994, p. 59). These findings can be interpreted as an indication of limited information processing related to anger. Moreover, Tiedens and Linton (2001) showed that people experiencing high certainty depend on an expertise source cue in their assessments more compared to people experiencing low certainty when they feel anger (or contentment).

However, other emotional theories have suggested quite opposite outcomes. The affect-as-information model (Schwarz, 1990) argues that affect reflects the environment's hospitability.

Positive affect indicates an innocuous environment and encourages minimal processing of cognitive resources, while negative affect indicates a hostile environment and activates demanding and thorough processing to deal with struggles. Anger similar to other negative emotions elicits analytic processing to tackle with difficulties. This finding is suggested by hedonic-contingency theory (Wegener & Petty, 1994), according to which people urge to make strategic processing choices to maintain their positive moods and improve negative moods. In this sense, three experimental studies by Moons and Mackie (2007) demonstrated that angry people are likely to process information analytically, that heuristic cues can influence them even as they do process analytically, and that they selectively use only relevant cues.

Consequently, because of the higher analytical processes underway when people write anger-embedded reviews, they might be more affected by prior average ratings and adjust their ratings more, which means reviewers' evaluations are more likely to be affected by anchoring after writing an anger-embedded review. In formal terms:

H2: The positive relationship between prior average rating and subsequent review ratings is stronger (more positive) when anger increases.

2.2.3 Social influence and hotel characteristics

When they write reviews, consumers are not only exposed to prior average ratings, but they also see other prominent hotel quality indicators, namely hotel category and brand name. Previous research in the hospitality industry has demonstrated that both prices and customer ratings increase with each additional hotel star (i.e., category) (Martin-Fuentes, 2016). Xu (2019) showed that hotel attributes have different effects on customer evaluations for hotels with different star levels and for independent versus chain hotels.

Hotel grading is a method of categorizing hotels based on a variety of factors (WTO, 2014), with particular attention to the sort of services supplied to consumers. The category of hotel stars is a well-known international hotel system that denotes a hotel's excellence when it has multiple stars (Abrate, Capriello & Fraquelli, 2011). Room prices of hotels with higher star levels generally high, and both higher prices and star levels boost customer expectancies (Choi & Mattila, 2004). Moreover, in comparison to a four-star or lower-rated hotel, a five-star hotel is expected to offer a higher level of service. Consequently, a hotel's categorization can serve as a guide for consumers, telling them what they can expect in terms of service and performance from that hotel (Viglia, Minazzi & Buhalis, 2016).

Additionally, price level (high vs low) and quality evaluations (Zeithaml, 1988) provided by both previous customer ratings and the hotel category (i.e., luxury vs budget) can affect customer expectations of the level of service quality offered by a hotel (Viglia, Minazzi & Buhalis, 2016), influencing, in turn, their information processing.

According to Dai, Chan and Magnier (2020), consumers rely less on consumer reviews for experiential purchases compared to those for material purchases because they believe reviews of experiences reflect less objective qualities than those of material goods. Experiential purchases are hedonic in nature whereas material goods are utilitarian. Since luxury hotel guests value more hedonic attributes and benefits, whereas budget hotel guests value more utilitarian attributes and benefits (Prebensen & Rosengren, 2016; Rhee &Yang, 2015), it is possible that budget hotel rating is more likely to be positively affected by social influence (i.e., prior average rating) than with luxury hotels, because review content and ratings for budget hotels might be perceived as more objective and more helpful.

Additionally, in the case of a budget hotel, consumers will make less costly decisions and therefore they are characterized by low expectations (Choi & Mattila, 2004), which implies the adoption of less systematic thinking and higher adoption of heuristics, which in turn will lead to higher social influence when they write reviews. Unlike the case of luxury hotels, for which consumers will make more costly decisions with consequent higher expectations (Radojevic, Stanisic, & Stanic 2015; Abrate, Quinton, & Pera, 2021), budget consumers will deliberate more systematically, leading to less social influence when they write reviews.

Therefore, we hypothesize that:

H3: The positive relationship between prior average rating and subsequent review ratings is stronger (more positive) for budget hotels.

H4: The positive relationship between prior average rating and subsequent review ratings is weaker (less positive) for luxury hotels.

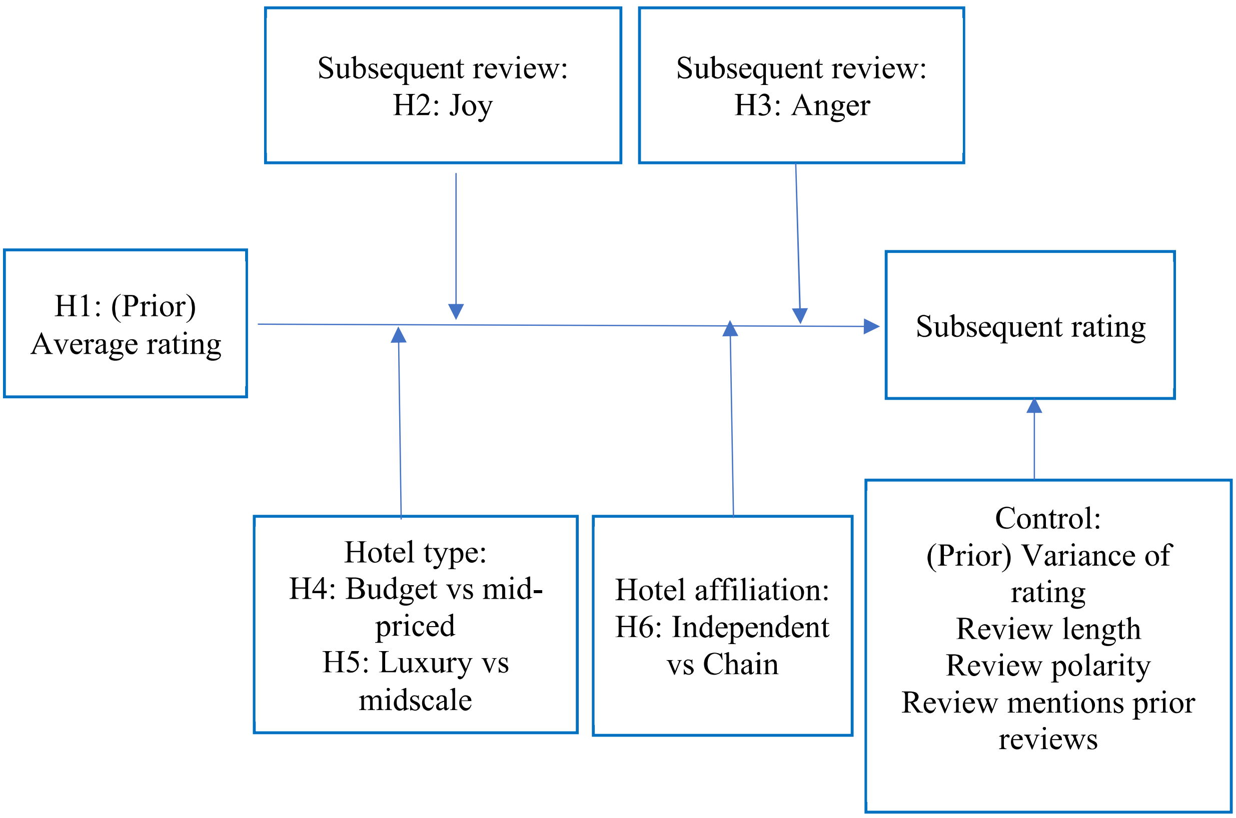
The importance of a brand in the hotel sector explains why industry participants have embraced brands and linked themselves with brand chains as a differentiating component of their marketing activities (Dev, Morgan & Shoemaker, 1995). Since independent hotels have small ownership structure, with consequent lack of sufficient resources, as implied by resource constraint theory, their performance can be poor (O'Neill & Carlback, 2011). Consequently, consumers might expect low performance for independent hotels that are not affiliated with a brand chain (Filieri, Vitari & Raguseo, 2021). Furthermore, compared to independent hotels, chain hotels are typically larger and contain more features, and they are therefore more likely to receive higher ratings (Banerjee & Chua, 2016).

Moreover, chain hotels likely have a very high number of reviews due to higher consumer loyalty compared to independent hotels (Kandampully & Suhartanto, 2003). Consequently, compared to independent hotels, travelers can perceive chain hotels as a popular option that conveys certainty regarding quality. They may therefore use others' ratings as an anchor and adjust their ratings accordingly. Thus, we expect that:

H5: The positive relationship between prior average rating and subsequent review ratings is stronger (more positive) for chain (vs independent) hotels.

The research framework was shown in fig. 2.1.

Figure 2.1. The Conceptual Framework



2.3 Method

2.3.1 Data

We collected hotel reviews from TripAdvisor, one of the most well-known online sources of hotel reviews (comScore, Inc, 2016), with over five million registered users who visit the site an average of 30 million times every month and with coverage of over a quarter of a million hotel ratings from all over the world (O'Connor, 2008). In addition, TripAdvisor is one of the hotel review websites most widely researched by the academic community.

We also extracted data covering the same period from Expedia to assess the robustness of our findings. Since TripAdvisor was part of the Expedia Group until 2011, comparison of the same hotel samples makes much more sense for the same time period.

The study sample included 92992 online reviews of 213 New York City hotels on TripAdvisor. These were later matched to the Expedia sample using hotel name and address matching, and finally we merged review data with information on hotel characteristics from the Olerey company. After omitting cases with missing values in the hotel price category, the final TripAdvisor sample covered 65215 hotel reviews from 169 hotels in New York City. We chose hotels in New York City, a leading tourism city that accommodates many domestic and international tourists annually, to include hotels with diverse price scales and to guarantee an adequate number of reviews for each hotel included. Hypotheses were tested using the TripAdvisor sample, and the robustness of findings was tested using the Expedia sample, which contained 67534 reviews of 169 hotels. Among these 169 hotels, the frequency of budget, economy, mid-priced, luxury, and upscale hotel categories were 16, 28, 68, 27, and 31, respectively.

Our final sample, which was in review/year observations, included reviews from hotels with all price ranges: budget (n = 8693 reviews for 16 hotels, 13.33 %), economy (n = 9756 reviews for 28 hotels, 14.96 %), mid-price (n = 25065 reviews for 68 hotels, 38.43 %), luxury (n = 9236 for 27 hotels, 14.16 %), and upscale (n = 12465 reviews for 31 hotels, 19.11 %). The distribution of reviews for the Expedia sample were 17% for budget (n = 11215), 16% for economy (n = 10826), 44% for mid-price (n = 29619), 11% for luxury (n = 7078), and 13% for upscale (n = 8796) hotels. Consumers' online hotel ratings, the review text, the hotel's class, and amenities, and voluntarily provided reviewer information were all included in the data.

2.3.2 Sentiment analysis of review texts

Sentiment analysis, or opinion mining, is the technique of employing text analysis to discover and categorize customer attitudes, thoughts, judgments, and emotions to extract consumers' opinions and sentiments in real time.

It is a non-intrusive method which helps to avoid possible recall biases from traditional self-report measurements (e.g., surveys or polls) and is more time-and cost-efficient (Geetha, Singha, & Sinha, 2017). We used a bag of words (BoW) model, meaning that looking at the words in the text, disregarding the word order and the grammatical details (Zhang et al., 2010). The algorithm extracts the document's words (a process known as tokenization), removes extraneous words (such as "a," "is,"). We used a hybrid approach by combining the lexicon approach Sentiment VADER (Valence Aware Dictionary and sEntiment Reasoner) in the Stanford Natural Language Toolkit (NLTK) in a Python environment with a machine learning algorithm Naïve Bayes classifier to calculate polarity measurement (Manning et al., 2014), an approach which has received a lot of attention in academia (e.g., Giatsoglou et al., 2017).

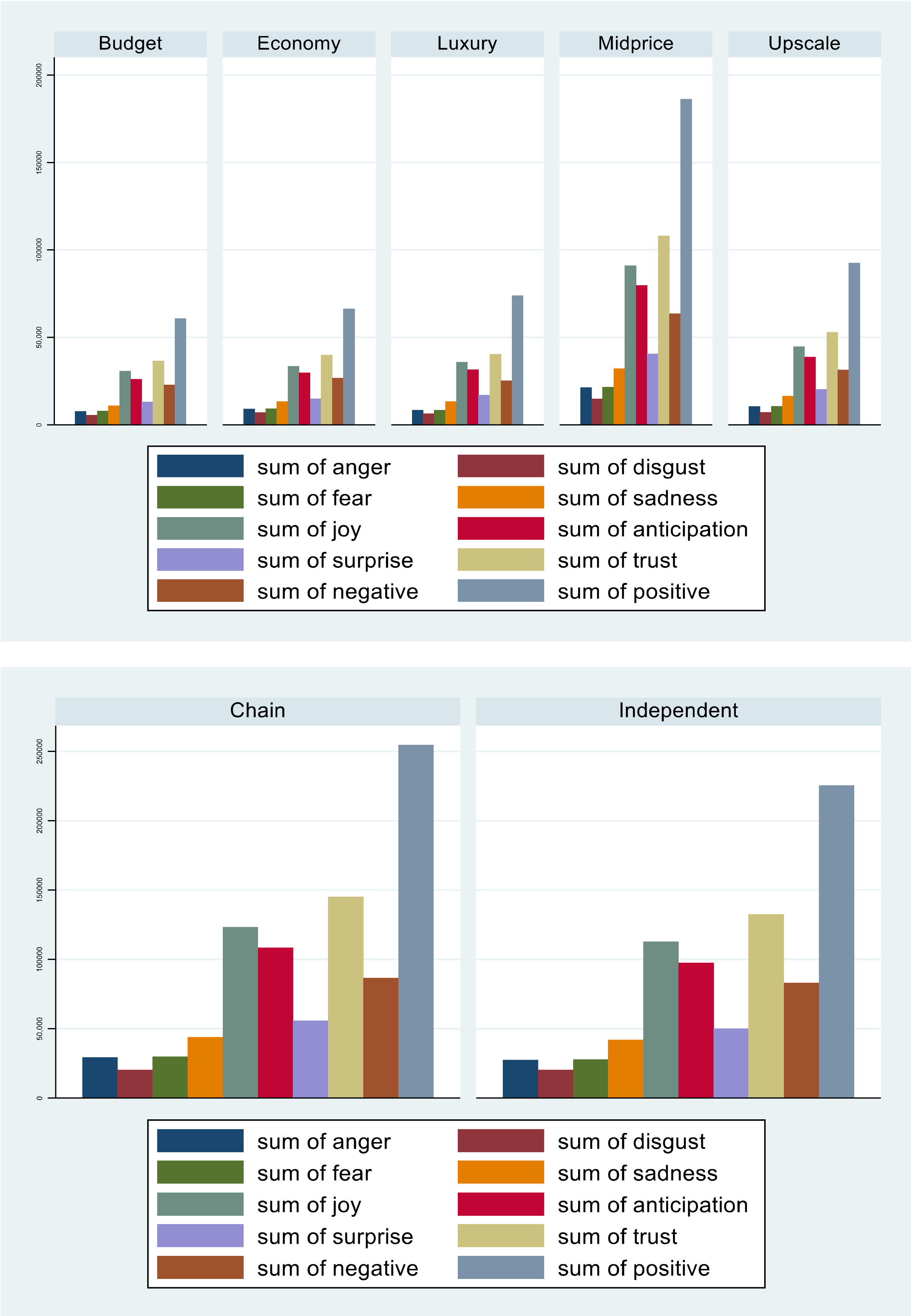
This feature assesses polarity intensity, with a value ranging from –1 (most extreme negative) to 1 (most extreme positive). VADER is a sentiment analysis lexicon which is a simple rule-based model for general sentiment analysis and has been proven to work exceptionally well for social media texts and online reviews vis-à-vis eleven typical state-of-practice benchmarks including polarity-based lexicons (the Linguistic Inquiry and Word Count (LIWC), the General Inquirer, and the Hu-Liu04 opinion lexicon), valence-based lexicons (the Affective Norms for English Words (ANEW), SentiWordNet, and SenticNet), context-aware lexicons (word-sense disambiguation (WSD)), and machine learning oriented techniques (Naive Bayes classifier, Maximum Entropy, and Support Vector Machine classification (SVM-C) and regression (SVM-R)) (Hutto & Gilbert, 2014).

2.3.3 Extracting emotions from a review text

Plutchik's multidimensional emotional framework was employed for three reasons. First, this emotional wheel is well supported by psychological research. Second, unlike some other alternative models (e.g., Ekman, 1992), Plutchik's framework balances positive and negative emotional perspectives. Third, it is a superset of certain other emotional dimensions (e.g., Ekman, 1992). Valence can be categorized as positive (joy and trust), neutral (surprise and anticipation), and negative (anger, disgust, sadness, and fear) (Laros & Steenkamp, 2005). There are two different levels of arousals: low arousal (trust, anticipation, sadness) and high arousal (joy, surprise, anger, disgust, and fear) (Reisenzein, 1994; Russell & Barrett, 1999). The NRC Word-Emotion Association Lexicon (EmoLex) (Mohammad et al., 2010) was selected to extract information on emotions because its dictionary size is large (14182 words) and rich, in terms of the emotional dimensions used.

Moreover, it has been already used and suggested by previous literature on emotional content (Felbermayr & Nanopoulos, 2016; Wang, Tang & Kim, 2019; Gang & Taeho, 2019). This lexicon is a collection of emotion words and their associated emotional intensities for various emotion categories. It divides words into eight basic emotional categories: “anger, anticipation, disgust, fear, joy, sadness, surprise, and trust,” as well as a higher-level category of “positive and negative.” First, the algorithm extracts the words that are associated with the eight emotions listed in Plutchik’s model of emotions. For each category of emotion, the algorithm counted the number of words from the lexicon that were present in the review. As explained earlier, we included two out of the eight emotions as the moderating variables in our study, namely joy and anger. Figure 2.2 indicates the sum of the eight emotion words in TripAdvisor reviews, according to hotel price and operational categories.

Figure 2.2 Sum of emotions words for hotel category and operation



2.3.4 Variables

The reviewer's rating (Y_{ijt}) on a scale of one (terrible) to five (excellent) stars was the dependent variable (Li et al., 2019, 2020). A series of review-text, reviewer-, hotel- and competitor-specific variables were included in the analyses.

Independent variables

Social influence was operationalized as the average of previous ratings of the hotel before the focal reviewer (Sridhar & Srinivasan, 2012, pg. 75). Thus, the independent variable "social influence" was measured as the prior average rating ($PriorAvgRating_{ij(t-1)}$) and it could have values between 1 and 5. Since TripAdvisor shows visitors the average prior ratings rounded to the nearest 0.5 points (i.e., bubble rating), we also rounded this value to the nearest 0.5 points[1].

The alternative measure of social influence is variance in prior review ratings ($VarPriorRating_{ij(t-1)}$) which was computed as the variance of other consumers' ratings before the focal reviewer (Li et al., 2020).

Moderators

Review texts were used to extract characteristics linked to consumers' product experience. The first review content-specific moderators are anger and joy, which were measured as the total number of emotional words associated with anger and joy in EmoLex.

Hotel-specific moderators are price scale and affiliation. Price scale was measured as a dummy variable, for five categories, namely budget, economy, luxury, mid-priced, and upscale, and the reference category was mid-priced. The most visible attribute of a hotel in the eyes of a customer is its affiliation.

[1] The results for Table 4 without the rounding of prior average ratings are available upon request from the authors. The only difference was that the moderating effect of the independent (vs chain) hotel variable on subsequent ratings was significant and negative.

A hotel might be unaffiliated ("independent") or a part of a branded chain. Affiliation was also measured as a dummy variable, indicating 1 for independent and 0 for chain hotel.

Control variables

We created a dummy variable called "neighbor hotel," which was coded as 1 if the hotel had a neighbor within 0.5 km to control competitor effects, consistent with Mayzlin et al., (2014). To do so we used the STATA Geodist tool to compute the distance between hotels in our sample based on their latitude and longitude. Since long reviews imply greater effort by the reviewer compared to short reviews and are more unfavorable than short reviews (Forman, Ghose & Wiesenfeld, 2008), we controlled for review length, which was assessed by the logarithm of the number of words in the review (Poncheri et al., 2008; Li et al., 2020). The natural logarithm transformation was used to normalize this measurement because the range of number of words was highly scattered.

The *polarity* score is ranges from -1 (most extreme negative) to 1 (most extreme positive), higher score implies that the text has a more positive sentiment.

A score of 0 indicates that the sentiment is neutral (Geetha et al., 2017). We additionally added a variable, reference to prior reviews (ReferPrior) indicating a dummy variable of whether review j for hotel i mentions a previous review (1) or not (0). If the review contains the words "review(s)" and "reviewer(s)", it indicates the review references a previous review (Hu & Li, 2011).

Table 2.2 describes the variables in our model. Table 2.3 illustrates correlations between the variables. The distribution of hotel ratings is very skewed – mostly 5 or 4, similar to previous findings (Mariani & Borghi, 2018). As is consistent with previous literature, the majority are either "5: very satisfied" 42 % or "4: satisfied" 34%, and a few are "3: so-so" 12%, "2: dissatisfied" 6%, and "1: very dissatisfied" 6% (Figure 2.3).

The average VIF of variables was equal to 1.51 and none of multicollinearity indicators of variance inflation factors (VIF) did exceed 10, the commonly accepted cut-off (Hair et al., 2006).

Figure 2.3 The frequency distribution of ratings (TripAdvisor)

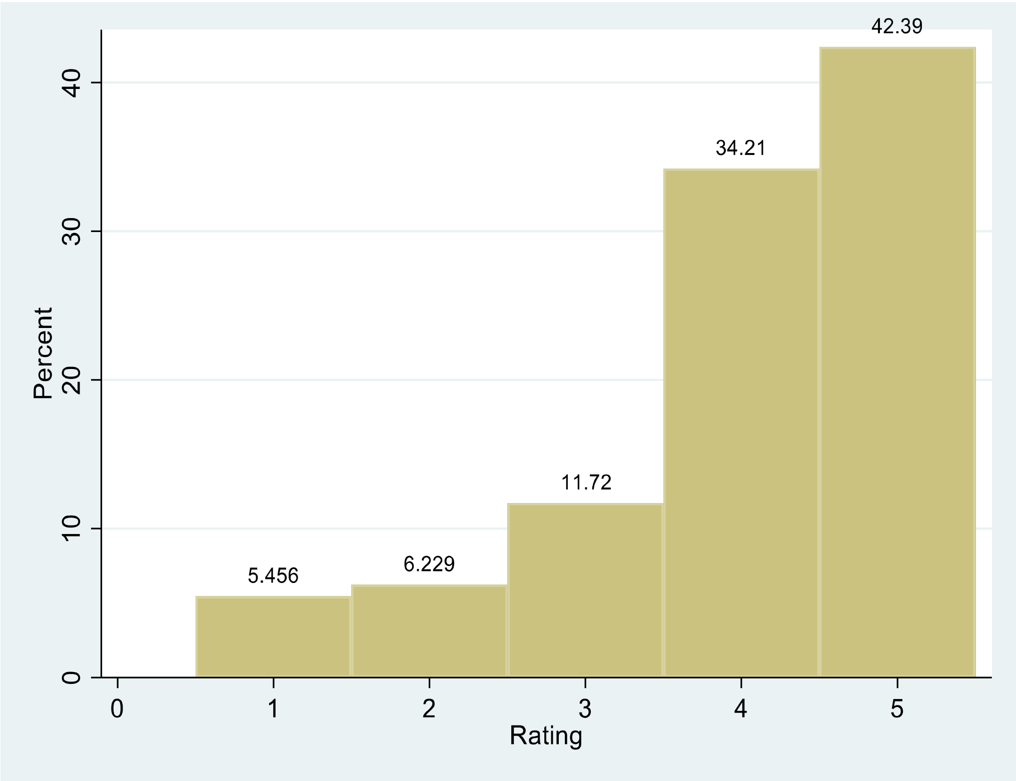


Table 2.2 Variable description

Variables	**Definition**
DV: Rating (Y_{ijt})	Review rating in review *i* for hotel *j* at time *t*
IV:PriorAvgRating$_{ij(t-1)}$	The prior average review rating for hotel *j* at time *t-1*
Moderators: **Review level** Anger$_{ijt}$	 Anger-embedded content in a review i for hotel j

Joy_{ijt}	Joy-embedded content in a review i for hotel j
Hotel level	
Price scale ($Price_{jt}$)	Price category of a hotel. Five hotel price category dummies: Budget, Economy, Mid-priced, Luxury and Upscale.
$Independent_{jt}$	Hotel affiliation category: 1: independent 0: chain
Control variables	
$VarPriorRating_{ij(t-1)}$	Variance of prior ratings
$Neighbor_{ij}$	Neighbor hotel dummy: 1 if the hotel had a neighbor within 0.5 km
	Logarithm of the number of words in review i.
$LogLength_{ijt}$	Sentiment score between -1 and 1 in a review text i.
$Polarity_{ijt}$	Dummy variable whether review i for hotel j refer to previous reviews
$ReferPrior_{ijt}$	(1) or not (0).

Table 2.3 Spearman correlation between variables (N = 65215)

Variables	Mean (SD)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1. Rating	4.02(1.13)	1.00							
2. PriorAvRating[a]	3.95(0.52)	0.36	1.00						
3.VarPriorRating[a]	1.15(0.47)	-0.28	-0.79	1.00					
4. Anger	0.87(1.23)	-0.23	-0.08	0.07	1.00				
5. Joy	3.62(2.86)	0.16	0.08	-0.06	0.33	1.00			
6. Budget	0.13(0.34)	-0.02	-0.03	-0.09	0.01	-0.01	1.00		
7. Economy	0.15(0.36)	-0.15	-0.30	0.16	0.02	-0.02	-0.16	1.00	
8. Mid-priced	0.14(0.35)	0.03	0.06	0.17	0.01	0.02	-0.16	-0.17	1.00
9. Luxury	0.38(0.49)	0.07	0.08	-0.13	-0.01	0.01	-0.31	-0.33	-0.32

10. Upscale	0.19(0.39)	0.06	0.16	-0.07	-0.02	0.00	-0.19	-0.20	-0.20
11. Independent	0.48(0.41)	-0.08	-0.23	0.19	0.01	0.00	0.31	0.25	-0.11
12. Neighbor	1.98(023)	-0.09	-0.19	0.12	0.01	-0.04	-0.30	0.02	0.03
13. LogLength	6.56(0.74)	-0.11	-0.02	0.03	0.49	0.56	-0.01	-0.01	0.01
14. Polarity	0.68(0.55)	0.38	0.16	-0.12	0.12	0.69	-0.01	-0.05	0.03
15. ReferPrior	0.14(0.35)	0.00	-0.03	0.01	0.18	0.22	0.03	0.03	-0.04

		(9)	(10)	(11)	(12)	(13)	(14)	(15)
9. Luxury	0.38(0.49)	1.00						
10. Upscale	0.19(0.39)	-0.38	1.00					
11. Independent	0.48(0.41)	-0.14	-0.22	1.00				
12. Neighbor	1.98(023)	0.05	0.16	-0.19	1.00			
13. LogLength	6.56(0.74)	0.00	0.00	-0.01	0.02	1.00		
14. Polarity	0.68(0.55)	0.03	0.00	-0.01	-0.06	0.41	1.00	
15. ReferPrior	0.14(0.35)	0.01	-0.04	0.05	-0.03	0.27	0.19	1.00

2.3.5 Econometric model

In this study, we explored two sets of moderating factors that influence subsequent ratings: review-content specific, namely joy- and anger-embedded reviews, and hotel-specific, namely the price scale category and ownership structure category. Additionally, we included reviewer-specific moderators (i.e., reviewer anonymity and expertness) to explain the social influence effect. Since there are multiple reviews per hotel and the dependent variable is ordinal in nature, a multilevel ordinal logit model was preferred. Standardized values of continuous variables were used in model estimation to make it possible to compare the magnitude of effects easily. The unit of observation is review per hotel in chronological order. The unit of analysis is the review.

A review rating y_{ijt} can take values among 1 and 5, which is the rating given by reviewer i for hotel j at time t. There is total 65215 reviews and 213 hotels. Y^*_{ijt} indicates the latent variable of reviewers' hotel evaluation. For instance, y^*_{ijt1} is the probability that hotel j is rated 1, and y^*_{ijt4} is the probability that hotel j is rated less than or equal to 4. The explanatory variables of reviewer hotel evaluations are presented in equation 1.

$$
\begin{aligned}
Y^*_{ijt} = \beta_0 &+ \beta_1 PriorAvgRating_{ij(t-1)} + \beta_2 Var(PriorRating)_{ij(t-1)} + \beta_3 Anger_{ijt} \\
&+ \beta_4 Joy_{ijt} + \beta_5 Budget_{jt} + \beta_6 Economy_{jt} + \beta_7 Luxury_{jt} + \beta_8 Upscale_{jt} \\
&+ \beta_9 Independent_{jt} + \beta_{10} REF_{ijt} + \beta_{11} PriorRating_{ijt} \\
&* Anger_{ijt} + \beta_{12} PriorRating_{ijt} * Joy_{ijt} + \beta_{13} PriorRating_{ijt} * Budget_{jt} \\
&+ \beta_{14} PriorRating_{ijt} * Luxury_{jt} + \beta_{15} PriorRating_{ijt} \\
&* Independent_{jt} + \beta_{16} PriorRating_{ijt} * ReferPrior_{ijt} + \beta_{17} Neighbor_{jt} \\
&+ \beta_{18} Polarity_{ijt} + \beta_{19} Loglength_{jt} + \beta_{20t} Year\ Dummies_{ijt} + \varepsilon_{ijt}
\end{aligned}
$$

2.3.6 **Results**

Table 2.5 presents the outcomes of the proposed research model. M0 is the model with only control variables. M1 is the base model, which includes the main effects of prior average rating, variance of prior ratings, anger, joy, price category dummies, hotel affiliation dummy, reviewer's expertise and anonymity, and control variables. M2 examined the interaction effects of the prior average rating with joy, anger, and the variance in prior ratings on subsequent review ratings. Model 3 investigated the additional price scale category interactions included in M2. M4 added hotel affiliation interaction and the budget and luxury hotel category interactions to base model M2. M5 included reviewer-specific interactions in M4.

Finally, Model 6 was the comprehensive model, including the moderating effects of the emotion embedded in reviews, variance of prior ratings, budget, and luxury hotel interactions, and whether a review mentioned previous reviews. Model 6 is thus utilized to describe the final estimation results below.

All control variables were significant in Model 0, except for having a neighbour hotel in close vicinity (b = -0.18, p = 0.28). Thus, the remaining models did not include this variable.

We first interpreted the main effects of prior average rating and moderating variables because they constituted the basis for all of our hypotheses. The main effects were coherent, indicating robustness in models 1-6. According to Model 1, prior average rating had a significant positive effect on subsequent ratings (b = 0.29, $p < 0.001$). Variance in prior ratings did not have a significant effect on subsequent ratings (b = -0.01, p = 0.52). Joy had a positive effect on subsequent ratings (b = 0.38, $p < 0.001$) and anger had a negative effect (b = -0.50, $p < 0.001$). Anger in reviews had a stronger effect on subsequent ratings than joy.

Ratings were lower for economy hotels (b = -0.30, $p < 0.01$) and budget hotels (b = -0.20, p = 0.12) compared to midscale hotels; however, interestingly, the effect was significant only for economy hotels. Luxury and upscale hotels had higher ratings than midscale hotels (b_{luxury} = 0.57, $p < 0.001$, $b_{upscale}$= 0.24, $p < 0.05$) and the effect was higher for luxury hotels. The ratings of independent hotels were lower than those for chain hotels (b = -0.06, p = 0.43), but it was not significant. Ratings were high when the review mentioned previous reviews or reviewers (b = 0.16, $p < 0.001$).

Hypothesis 1 investigated whether the positive effect of social influence on subsequent ratings was reduced when the review was embedded with joy.

According to Model 6, this hypothesis is supported because the interaction effect is significant and negative (b = -0.04, $p < 0.001$). Thus, the positive effect of social influence on subsequent ratings was weaker for joy-embedded reviews. When both prior average ratings and joy increased, subsequent ratings decreased.

Hypothesis 2 investigated whether anger-embedded reviews increased the impact of social influence on subsequent ratings. The effect of prior average rating was positively moderated by anger-embedded reviews (b = 0.02, $p < 0.05$); thus, an increase in anger increased the significance effect of the prior average rating. The social influence that the prior average rating had on subsequent ratings was stronger when anger increased in a review text. H2 was supported.

Hypothesis 3-4 stated that the social influence effect is weaker for luxury hotels and stronger for budget hotels. According to H3, the results in Model 6 indicated that the interaction effect produced by prior average ratings for budget hotels on subsequent ratings was significant and positive (b = 0.16, $p < 0.001$), but not for luxury hotels (b = -0.05, p = 0.29). Therefore, H4 was not supported.

Hypothesis 5 predicted a stronger relationship between prior average ratings and subsequent ratings when the hotel is independent. The results in Model 6 indicated that the interaction effect produced by prior average ratings for independent hotels on subsequent ratings was not significant (b = 0.00, p = 0.90). Therefore, H5 was not supported.

Additionally, we checked whether the social influence effect changes with variance in prior ratings and when a review mentions prior reviews or reviewers. Variance of prior ratings significantly moderated the positive effect of prior ratings on subsequent ratings (b = -0.03, $p < 0.05$).

Thus, the positive influence of prior average ratings was less positive with increasing variance in prior ratings. The positive influence that prior ratings had on subsequent ratings was weaker (less positive) when the review mentioned previous reviews or reviewers (b = -0.12, $p < 0.01$).

The impacts of control variables on consumer hotel evaluations were consistent and similar. Review length had a significant and negative effect on ratings (b = -0.28, $p < 0.001$), indicating dissatisfied hotel guests wrote longer reviews. Polarity had a significant positive influence on ratings (b = 0.62, $p < 0.001$); that is, increased positive sentiment increased ratings. When the review referred to prior reviews or reviewers, subsequent ratings increased (b = 0.13, $p < 0.001$).

Time effects indicated that hotel ratings increased significantly at TripAdvisor over time. Specifically, there was a significant increase in ratings when we tested the time coefficient of 2011 to 2005 ($chi2(1) = 180.54$, $p < 0.001$).

Table 2.4 Summary of Hypotheses-Testing Results

Hypothesis	Empirical support
H1: The positive relationship between prior average rating and subsequent review ratings is weaker (less positive) when joy increases.	√
H2: The positive relationship between prior average rating and subsequent review ratings is stronger (more positive) when anger increases.	√
H3: The positive relationship between prior average rating and subsequent review ratings is stronger (more positive) for budget hotels.	√
H4: The positive relationship between prior average rating and subsequent review ratings is weaker (less positive) for luxury hotels.	X
H5: The positive relationship between prior average rating and subsequent review ratings is stronger (more positive) for chain (vs independent) hotels.	X

Table 2.5 TripAdvisor (standardized values used)

	M0	M1	M2	M3	M4	M5	M6
LogLength	-0.28***	-0.28***	-0.28***	-0.28***	-0.28***	-0.28***	-0.28***
Polarity	0.63***	0.63***	0.63***	0.63***	0.63***	0.63***	0.62***
ReferPrior	0.15***	0.16***	0.15***	0.15***	0.15***	0.15***	0.13***
Anger	-0.50***	-0.50***	-0.50***	-0.50***	-0.50***	-0.50***	-0.50***
Joy	0.38***	0.38***	0.38***	0.38***	0.38***	0.38***	0.38***
Budget	-0.39*	-0.20	-0.18	-0.08	-0.07	-0.07	-0.07
Economy	-0.41**	-0.30**	-0.29**	-0.35***	-0.31**	-0.31**	-0.31**
Luxury	0.78***	0.57***	0.58***	0.60***	0.60***	0.60***	0.60***
Upscale	0.34*	0.24*	0.24*	0.23*	0.24**	0.24**	0.24**
Independent	-0.21	-0.06	-0.05	-0.03	-0.03	-0.03	-0.03
Neighbor hotel	-0.18						
Prior Average Rating		0.29***	0.29***	0.29***	0.28***	0.28***	0.30***
Variance prior ratings		-0.01	-0.03	-0.04*	-0.04*	-0.04*	-0.04*
PriorAvgRating×Anger			0.02	0.03	0.02	0.02	0.02*
PriorAvgRating×Joy			-0.05***	-0.05***	-0.05***	-0.05***	-0.04***
PriorAvgRating×Variance			-0.03*	-0.02*	-0.03*	-0.03*	-0.03*
PriorAvgRating×Budget				0.14**	0.16**	0.16**	0.16***
PriorAvgRating×Economy				-0.08			
PriorAvgRating×Luxury				-0.07	-0.05	-0.05	-0.06
PriorAvgRating×Upscale				0.01			
PriorAvRating×Independent					0.00	0.00	0.00
PriorAvgRating×ReferPrior							-0.12***
2005	0.02	0.04	0.04	0.04	0.04	0.04	0.05

2006	0.13**	0.15***	0.16***	0.16***	0.16***	0.16***	0.16***
2007	0.16***	0.19***	0.20***	0.19***	0.20***	0.20***	0.20***
2008	0.28***	0.30***	0.31***	0.30***	0.30***	0.30***	0.31***
2009	0.45***	0.44***	0.46***	0.45***	0.45***	0.45***	0.46***
2010	0.54***	0.51***	0.52***	0.52***	0.52***	0.52***	0.52***
2011	0.69***	0.64***	0.65***	0.65***	0.65***	0.65***	0.65***
cut1	-3.64***	-3.25***	-3.22***	-3.22***	-3.22***	-3.22***	-3.23***
cut2	-2.53***	-2.14***	-2.11***	-2.11***	-2.11***	-2.11***	-2.11***
cut3	-1.39***	-0.99***	-0.97***	-0.97***	-0.97***	-0.97***	-0.97***
cut4	.58	0.98***	1.01***	1.01***	1.01***	1.01***	1.01***
var(cons)	.42***	0.18***	0.16***	0.15***	0.15***	0.15***	0.15***
N	65046	65046	65046	65046	65046	65046	65046

2.3.7 Robustness Check: Comparison of Social Influence in TripAdvisor and Expedia Reviews

The robustness of our findings was inspected using online reviews extracted from Expedia for the same hotels in the TripAdvisor data for the same time period. One of the differences between the two platforms is that only verified guests can leave a review on Expedia. While ratings on Expedia were higher than those on TripAdvisor (4.38 vs 4.02), reviews on TripAdvisor had more emotional words which were heavily positive emotions (7.36 vs 3.05), and longer (6.56 vs 5.46) and diverse opinions (i.e., high variance, 0.47 vs 0.29), compared to those on Expedia (Table 2.6). Interestingly, differences between 4 (very good) and 5 (excellent) ratings in five-point rating systems was not significant in the Expedia sample; however, it was significant in the TripAdvisor sample, indicating more discriminatory evaluations in reviews on TripAdvisor compared to Expedia.

In terms of the social influence effect and moderators, the majority of our findings were consistent with those on TripAdvisor (Table 2.7). Similar to TripAdvisor, the association between previous average ratings and consequent review ratings in Expedia was also positively moderated by anger and the budget hotel category, and negatively by joy. The influence of prior average ratings in Expedia was not moderated by variance in prior ratings for Expedia. This might be due to low variation in rating evaluations in reviews on Expedia (i.e., low variance in prior ratings for Expedia reviews). Interestingly, differing from TripAdvisor data, the relationship between prior ratings and subsequent ratings did not change when the review text mentioned prior reviews or reviewers in the Expedia data. Since TripAdvisor is more than a booking platform, meaning a platform for not only booking but also for looking for information about travel sites, consumers might think that the audience is bigger in size and social presentation concerns might play an important role in reviewer rating decisions. That is why the social influence effect might be higher in TripAdvisor reviews compared to Expedia ones. Indeed, the main effect of prior average review ratings on ratings in the TripAdvisor data was higher than the one in the Expedia one (0.29 vs 0.20).

Table 2.6 Summary statistics for TripAdvisor and Expedia

	TripAdvisor					Expedia			
	M	SD	Min	Max		M	SD	Min	Max
Rating	4.02	1.13	1	5		4.38	0.86	1	5
Prior average rating	3.96	0.51	1	5		4.37	0.34	1	5
Prior average rating (rounded)	3.95	0.52	1	5		4.38	0.35	1	5
VariancePriorRating	1.15	0.47	0	8		0.64	0.29	0	8
Anger	0.87	1.23	0	16		0.30	0.64	0	8

Joy	3.62	2.86	0	31		1.67	1.76	0	17
Anticipation	3.16	2.77	0	27		1.30	1.52	0	14
Disgust	0.62	1.07	0	16		0.20	0.52	0	7
Fear	0.89	1.26	0	20		0.34	0.69	0	9
Sadness	1.32	1.56	0	18		0.48	0.80	0	11
Surprise	1.62	1.65	0	18		0.64	0.92	0	9
Trust	4.26	3.33	0	36		1.91	1.95	0	16
Negative emotions	2.60	2.80	0	37		0.96	1.33	0	16
Positive emotions	7.36	5.69	0	68		3.05	3.01	0	27
LogLength	6.56	0.74	1.79	9.32		5.46	1.10	0	8.24
Polarity	0.68	0.55	-1	1		0.52	0.55	-1	1
REF	0.14	0.35	0	1		0.02	0.15	0	1

Table 2.7 Number of reviews that include anger and joy in TripAdvisor.

Anger	**Freq.**	**Percent**	**Cum.**
0	**33,446**	**51.29**	**51.29**
1	**31,769**	**48.71**	**100.00**
Total	**65,215**	**100.0**	

Joy	**Freq.**	**Percent**	**Cum.**
0	**8,603**	**13.19**	**13.19**
1	**56,612**	**86.81**	**100.00**
Total	**65,215**	**100.00**	

Table 2.7 shows how frequently and how many of the reviews are rated as "anger review" or "joy review". For Anger, there were 65,215 reviews in total, with 33,446 (51.29%) classified as not expressing anger (0) and 31,769 (48.71%) classified as expressing anger (1).

For Joy, there were also 65,215 reviews in total, with 8,603 (13.19%) classified as not expressing joy (0) and 56,612 (86.81%) classified as expressing joy (1).

Table 2.8 Results of multilevel ordered logit regression (standardized values used) for Expedia

	M0	M1	M2	M3	M4	M5	M6
LogLength	-0.04***	-0.04***	-0.04***	-0.04***	-0.04***	-0.04***	-0.04***
Polarity	0.45***	0.45***	0.45***	0.45***	0.45***	0.45***	0.45***
ReferPrior	0.19***	0.20***	0.19***	0.19***	0.19***	0.19***	0.20***
Anger	-0.31***	-0.31***	-0.30***	-0.30***	-0.30***	-0.30***	-0.30***
Joy	0.15***	0.15***	0.14***	0.15***	0.15***	0.15***	0.15***
Budget	-0.59**	-0.36*	-0.35*	-0.25	-0.25	-0.25	-0.25
Economy	-0.65***	-0.49***	-0.49***	-0.50***	-0.54***	-0.54***	-0.54***
Luxury	0.74***	0.62***	0.62***	0.61***	0.62***	0.62***	0.62***
Upscale	0.22	0.19	0.19	0.19	0.19	0.19	0.19
Independent	-0.28*	-0.22*	-0.22*	-0.23*	-0.23*	-0.23*	-0.23*
Neighbor	-.14						
Prior Average Rating		0.20***	0.20***	0.10**	0.12***	0.12***	0.12***
Variance prior ratings		0.01	0.02	-0.01	0.00	0.00	0.00
2005	0.35**	0.35**	0.34**	0.36**	0.35**	0.35**	0.35**
2006	0.51**	0.43***	0.42***	0.44***	0.43***	0.43***	0.43***
2007	0.69***	0.58***	0.57***	0.59***	0.58***	0.58***	0.58***
2008	0.84***	0.71***	0.70***	0.72***	0.71***	0.71***	0.71***
2009	0.91***	0.78***	0.77***	0.79***	0.78***	0.78***	0.78***
2010	0.88***	0.74***	0.73***	0.74***	0.74***	0.74***	0.74***
2011	0.67***	0.54***	0.53***	0.54***	0.53***	0.53***	0.53***

PriorAvgRating×Anger			0.02*	0.02*	0.02*	0.02*	0.02*
PriorAvgRating×Joy			-0.04***	-0.04***	-0.04***	-0.04***	-0.04***
PriorAvgRating×Variance			0.01	0.00	0.00	0.00	0.00
PriorAvgRating×Budget				0.23***	0.18***	0.18***	0.18***
PriorAvgRating×Economy				0.11*			
PriorAvgRating×Luxury				0.05	0.02	0.02	0.02
PriorAvgRating×Upscale				-0.01			
PriorAvRating×Independent					0.03	0.03	0.03
PriorAvgRating×ReferPrior							0.03
cut1	-4.87***	-4.51***	-4.53***	-4.54***	-4.54***	-4.54***	-4.54***
cut2	-3.54***	-3.17***	-3.19***	-3.19***	-3.20***	-3.20***	-3.20***
cut3	-2.28***	-1.90***	-1.92***	-1.93***	-1.93***	-1.93***	-1.93***
cut4	-.24	0.13	0.12	0.11	0.11	0.11	0.11
var(cons)	0.48***	0.28***	0.29***	0.27***	0.27***	0.28***	0.28***
N	67365	67365	67365	67365	67365	67365	67365

2.4 Discussion

The objective of this study was to enhance the understanding on what conditions may change the impact that social influence, by way of prior average ratings, has on subsequent hotel review ratings. As they are moderators of social influence, we considered the emotions embedded in review texts indicating customer experience, hotel affiliation and hotel price scale categories. This paper employed text analytics to extract emotions from review texts. Our model considered the clustering nature of data, and thus dependence between observations (reviews) for the same hotel.

Moreover, we included a broad variety of meaningful control variables, such as the competitor effect and whether a review mentioned previous reviews, along with polarity and length. Our findings show that when prior average ratings increased, subsequent ratings increased as well, confirming previous studies (Ma et al., 2014; Guo & Zhou, 2016; Li et al., 2019; Li et al., 2020;). Thus, social influence effect, in terms of prior average ratings, was also proven to be positive in a hotel context. Hu and Li (2011) showed the negative effect of social influence on subsequent ratings for book reviews. Since books are related to identity and taste, reviewers might prefer more differentiating behaviors. The main effect that dispersion in prior ratings had on subsequent ratings was not significant, which was similar to the findings of Hu and Li (2011) and Lee, Hosanagar, and Tan (2015), but it was different from the previous findings of Guo and Zhou (2016), and of Li, Meng, Jeong, and Zhang (2020).

The influence that prior average review ratings had on subsequent ratings was stronger when the emotion expressed in reviews was anger, whereas this influence was weaker when a customer wrote a joy-embedded review. Thus, after a joyful consumer experience, when consumers see high prior average ratings, they reduce their ratings, which might cause a decreasing trend over time for a hotel. However, reviews with anger emotions might adjust and increase their subsequent ratings due to a reduction in their negative feelings and in order to feel better. These findings are coherent with mobilization-minimization theory (Taylor, 1991), according to which consumers respond to negative emotions with an initial phase of rejection (i.e., mobilization) of the stimulus that generated the negative emotion followed by a subsequent phase of cognitive effort aimed at reducing the undesired effects of the negative emotion (i.e., minimization). Both mobilization and the subsequent phase of cognitive effort is not required for positive emotions, as people do not need to cope or reduce any negative effect.

Consequently, positive emotions are generally characterized by constant responses over time, whereas negative emotions generate an initial instinctive reaction, followed by a more deliberative response. Donato and Miceli (2020) proved the mobilization-minimization theory for disgust, showing an initial response of closure followed by a subsequent response of openness. Similarly, the present research has demonstrated that anger regarding a negative service experience after an initial negative response (i.e., mobilization) in review writing is followed by an opposite response that minimizes, and even wipes away the negative product experience though a higher rating adjustment. Conversely joy regarding a positive experience does not generate rating adjustments.

While the effect of prior average ratings on consequent ratings was stronger for budget hotels, the effect that social influence had on subsequent ratings did not change for luxury hotel guests and independent hotels. Consistent with previous research (Li et al., 2020), the positive effect of prior average ratings was less positive with increasing variance in prior ratings. The positive influence that prior ratings had on subsequent ratings was weaker (less positive) when the review mentioned previous reviews or reviewers (Hu & Li, 2011).

Additionally, when review length increased, subsequent ratings decreased. Polarity also had a positive effect on subsequent ratings. The findings from TripAdvisor reviews were compared to those on Expedia for the same hotel sample. Similar to TripAdvisor data, the moderating effects of joy, anger and budget hotels were significant, whereas the effects of luxury and independent hotels were not significant in Expedia. However, unlike TripAdvisor, when the review text on Expedia referred to previous reviews or reviewers, the effect that social influence had on ratings did not change.

The social influence effect was found to be high in magnitude in TripAdvisor reviews compared to those on Expedia. Expedia ratings were higher than TripAdvisor ratings, in line with previous findings (Mayzlin et al., 2014), but were different from Xiang et al. (2017). Furthermore, TripAdvisor reviews had more emotional words, with the majority being positive ones. TripAdvisor reviews were longer and indicated diverse opinions compared to Expedia ones. Since there is no requirement of being a hotel guest to post on TripAdvisor and only a customer can post a review on Expedia, promotional reviews might happen more frequently in TripAdvisor, confirming the results of Mayzlin et al. (2014).

2.4.1 Scientific implications

There are several contributions of this research. Firstly, it is one of the few studies in hospitality and tourism literature that shows how social influence affects rating behaviors in hotel reviews. Online reviewers, who act as opinion leaders and influence others, may be socially affected. Consumers' online reviews and ratings are influenced by others' ratings. More importantly, the effect that social influence has on subsequent ratings will change depending on the emotions expressed and the hotel's characteristics.

Secondly, this is one of the first studies of online review platforms to look at the impact of previous ratings on subsequent review ratings using reviews with different embedded emotions (i.e., anger and joy). Thirdly, this is the first study to use a text mining approach to assess the moderating influence of emotions embedded in review texts and product characteristics. Emotions embedded in reviews have been used mostly to explain review helpfulness (Yin et al., 2014; Felbermayr & Nanopoulos, 2016; Ren & Hong, 2019; Chen & Farn, 2020), but not individual subsequent ratings.

By investigating hotel characteristics and two of the most important emotions in reviews, the current study also contributes to social influence studies in terms of prior average ratings (Ma et al., 2014; Schlosser, 2005; Li & Hitt 2008; Moe & Schweidel, 2012; Lee, Hosanagar & Tan, 2015; Ho, Wu & Tan, 2017). This study also supports previous findings on the bidirectional effect that social influence has on subsequent ratings (Sridhar & Srinivasan, 2012; Moe & Trusov, 2011; Munchik et al., 2013), in contrast to extant research indicating only the negative effect of social influence (Schlosser, 2005).

2.4.2 Managerial implications

While the main goal of reputation platforms is to persuade prospective customers to purchase a product or a service thanks to customers reviews, social influence – specifically through prior average ratings – is not always advantageous for a company, because a customer's experience of emotions or product characteristics might cause some customers to alter their ratings and not give an honest rating after being exposed to prior ratings by others. A better understanding of the moderators that can reduce or increase social influence would benefit the industry in terms of accurate product evaluations; thus, this study has investigated the role of embedded emotions as a proxy for guest experience indicators and product characteristics in online consumer hotel reviews.

Our findings have significant management implications for hotel managers, for the managers of reputation platforms, and for the designers of online rating systems.

Our findings can be used by reputation platform managers to adopt new forms of summarizing ratings from prior reviews or to prevent exposing reviewers to average ratings during review writing.

Budget hotels benefit more from social influence compared to luxury and chain hotels. Indeed, in their case, exposing consumers to past prior ratings is good. Luxury hotels do not need to worry much about the social influence effect because the positive effect of social influence does not change for them. Some de-biasing adjustments in the presentation of prior average ratings by hotel price type could be done on reputation platforms. Another form of de-biasing could be to show both a hotel's prior average rating and the average prior ratings of that hotel type. Reviews which mention previous reviews or reviewers can be used as an easy way to de-bias prior average ratings because they indicate reviews with high social influence.

2.4.3 Limitations and future research

There are several limitations to this research. Firstly, despite eight emotions being extracted from reviews, only two emotions were used in our model due to high correlations between emotion dimensions. Despite anger and joy being the most frequent emotions found in review texts, there is a need to investigate additional discrete emotions that can be found embedded in reviews (e.g., sadness, fear, surprise).

We did not consider selection bias in reviews or the problem of fake reviews. Consumers select themselves to write reviews and hotel guests who do not leave a review might be different than guests who write reviews. According to Han and Anderson (2020), only consumers who have extreme opinions write reviews and give ratings. We have also acknowledged the possibility of fake reviews being present on TripAdvisor (O'Connor, 2010). Moreover, while our research model considered many crucial elements related to social influence in online reviews, it omitted some information regarding reviewer characteristics.

Future research can investigate the effects these factors have, including reviewer nationality and gender. Finally, the data set was gathered from TripAdvisor (New York hotels) for the period between 2004 and 2011, and it is necessary to verify how these findings may be applied to different countries and more recent periods. Cross-cultural study on the moderators of social influence effect is possible avenue for future research.

CHAPTER 3

A longitudinal analysis to monitor changing patterns of hotel customers' perceptions using Topic modelling and sentiment analysis

3.1 Introduction

The hospitality industry has experienced a significant shift towards online bookings, with many customers sharing their feedback and reviews after their stay (Mehta et al., 2023; Song et al., 2022; Putranto et al., 2021; Chatterjee, 2020; Zhao et al., 2019). These reviews can take the form of either textual feedback or numerical ratings and have a significant impact on prospective customers' booking decisions through electronic word of mouth (eWOM) effects. The internet has contributed to increase the interconnection between the consumer communities in the specific context of hotels generating a high volume of online hotel reviews which makes it difficult to analyze (Chatterjee, 2020). The different comments on social platforms represent an important source of information for both (hotel consumers and hotel managers). Therefore, this study uses online reviews to better understand the factors that influence consumers' satisfaction and preferences. As a consequence, the goal of this study is to define the important aspects for choosing hotels based on consumers' opinions on TripAdvisor platform.

Previously, Ye et al. (2011) concluded that a large percentage of hotel customers trust on the online user-generated reviews to make online purchase decisions for hotels, higher than any other product category in the service industry.

Thus, more recently, several authors indicate the importance for hotel managers to effectively analyze their online customer reviews to better understand their customers, improve their performance and be more competitive (Mehta et al., 2023; Glavely et al., 2022; Phillips et al., 2020; Berenzina et al., 2016).

Furthermore, hotel segment is a crucial factor to consider by previous studies as it can lead to a deeper understanding of customers' purchasing behavior in different marketplaces in the tourism industry (Brito et al., 2015). According to Talluri and Van Ryzin (2005), hotels are categorized as city/business, extended-stay, resort or as a mix and can also be categorized by size and location. However, only a few numbers of studies have tried to analyze how customers' online rating patterns in the hotel review websites differ between the different categories of hotels (budget/economy and premium/luxury) and across different regions. For instance, within the context of Indian hotels, Geetha et al. (2017) explored the consistency of the relationship between customer review sentiments and hotel ratings across budget and premium hotels. Other authors collected review data from city and resort hotels in Portugal (Antonio et al., 2017) to obtain a prediction model for review ratings. In the context of China, Tian et al., (2016) examined English written online reviews in three-to-five-star hotels in four big cities and in the same country, more recently, Song et al. (2022) studied the differences in hotel customer satisfaction in Chengdu using online review data during the pandemic (from February to May 2020) and compared the results with the reviews for the same period in 2019. Moreover, in the context of Greece, Glaveli et al. (2022) made an analysis on hotel quality management and customer satisfaction from online ratings extracted from TripAdvisor for the overall population of hotels operating in a very popular destination, the Ionian Sea islands.

Very recently, Mehta et al. (2023) conducted an analysis in four and five star-rated hotels and resorts in popular travel developing countries like India, Indonesia, Malaysia, Singapore and Sri Lanka. They extracted travelers' reviews from Tripadvisor to study hotel customers' expectations during the first stages of the pandemic COVID-19 in 2020. In order to fill the gap in the literature addressed by Padma and Ahn (2020) who considered that previous studies gave little attention on service quality and customer satisfaction in luxury hotels in developing countries. Also, to address this issue, Putranto et al. (2021) revealed the most relevant topics through a longitudinal analysis in the period between 2015 and 2019 considering customer review data posted in English collected from three to five-star hotels in Indonesia.

Currently, as far as the authors know, there has only been developed two studies using customers' online opinions in the Spanish hotel context. One of them (Fuentes Medina et al., 2017) is case study research focused on The Spanish Tourist Paradors chain (emblematic hotels) through the analysis of content of online opinions of guests posted on the establishment's website. However, they did not go deeper determining the positive or negative meaning of the comments for each item of the value chain and they made a call for future research to address this limitation. The second study by Ahani et al. (2019) examined the online reviews from hotel customers in Spain and developed a new method for the use of multi-Criteria Decision-Making and soft computing approaches to reveal customers' preferences and satisfaction through a segmentation analysis. This second research used the case of Canary Islands hotels but only took into consideration customers' ratings and did not consider the valuable data hidden in the online review's texts and therefore they made a call for future studies in order to explore the impact of customers' online reviews and textual comments on customer satisfaction more accurately.

Despite the fact that online reviews and ratings have become crucial for travellers' decision-making, there are currently a few studies that demonstrate the effectiveness of online ratings on consumers' decision-making and travellers' preferences and satisfaction from the hotels' rating (Gavilan et al., 2018; Yu et al., 2017; Zhao et al., 2019). Due to the high competition in the tourism sector, customer satisfaction must be emphasised by hotels as a crucial factor in order to foster loyalty and positive attitude (Cetin and Dincer, 2014). Therefore, the need to use sentiment analysis and topic modeling methods is rapidly growing as a new research technique that used to analyse consumer opinions (Ban et al., 2019). Recent research has actively used topic modeling to uncover the hidden meanings and topics from the text (Kwon et al., 2021) while other studies have used sentiment analysis which uses natural language processing (NLP) to better understand whether people have negative or positive opinion posted in reviews (Liu et al., 2012).

To better understand how ratings behaviour is affected by positive and negative online reviews and to explore the perceived service quality attributes in hotels by predicting the variation of customer preferences in luxury hotel users in Spain, our research fills a gap in the literature by using sentiment analysis and topic modeling to discover what customers are satisfied with and dissatisfied with and to explain how we can monitor changing patterns of hotel customers' perceptions. Additionally, we can enhance the qualitative interpretation of textual data by using the topic modeling from luxury hotel reviews. Therefore, this study establishes some research questions to guide our empirical analysis:

1) How positive and negative online reviews influence ratings behaviours in luxury hotels?

2) What type of topics were discussed in consumer for luxury hotel reviews?

3) Are there any effect of topics and emotions on satisfaction ratings for luxury hotel reviews?

Therefore, the goal of this study is to extract topics from related words to understand what customers on luxury hotels talk about and how these topics explain customer satisfaction, as well as to understand the opinions and feelings of consumers expressed in online reviews toward the hotel services by analysing the impact of customers' emotions on review ratings. Our study also aims to explain how changing patterns of hotel customers' perceptions can be monitored through a longitudinal analysis during 17 years before the COVID-19 pandemic. This paper draws meaningful conclusions that offer hotel managers a chance to get help with their decision-making process and to provide them with crucial data and information that are useful for their management.

This paper is organized as follows. First, we develop a literature review and explain the expectancy confirmation theory (ECT) for understanding customer satisfaction. Secondly, we examine the determinants of online hotel customer satisfaction through studies that consider the technical attributes implied by the online hotel textual reviews and ratings. Thirdly, we assess travellers' evaluations of service attributes through some techniques for textual data analysis based on sentiment analysis and topic modelling. Specifically, we explain a topic model that has been applied in different contexts in the service quality literature like Latent Dirichlet Allocation (LDA). Three stages of topic modelling techniques are identified through our empirical research using TripAdvisor hotel reviews. We started by gathering the data needed for topic modelling in the last 17 years (2002-2019). The second stage involved pre-processing analysis to turn unstructured data into data suitable for topic modelling, and the final stage involved data analytics. Next, we provide the results and develop a discussion chapter. Finally, conclusions, theoretical and practical implications, limitations, and recommendations for future research are established.

3.2 Literature Review

3.2.1 The expectancy-confirmation theory (ECT) and the Signaling theory:

The expectancy-confirmation theory was introduced by Oliver (1980) and has been used for understanding customer satisfaction. According to this theory, if customers' perceived quality is higher than their expectation, customers are satisfied, and, conversely, if a service experience does not match or exceed customers' expectations, they will feel dissatisfied. For hotel customers, one way to show their satisfaction/dissatisfaction is through their evaluations of products or services through online customer ratings. According to expectation-confirmation theory, customer satisfaction is a function of customer evaluations that steams from a comparison (confirmation) between their pre-purchase expectations and their perceived quality of products and services after consumption, and therefore, in online reviews, customers mention their pre-expectations of quality of products and services, their perceptions, or both in order to describe why they are satisfied or dissatisfied (Zhao et al., 2019).

The other theory that provides a theoretical foundation for this study is the Signaling Theory. It is based on the existence of information asymmetry between parties and describes the signal behaviour between two parties (Connelly et al., 2011). In the hotel context, many products and services offered are intangibles and this characteristic means that there is an asymmetry of information about the quality of products and services between these two parties (hotels and customers). Hence, when hotel customers write online reviews after their experience, the contents and linguistics characteristics of their reviews are signals about their perceptions to hotel managers and prospective customers (Zhao et al., 2019; Geetha et al., 2017).

Even more, what customers write (their contents) and how they write (their linguistic style) signal their satisfaction or dissatisfaction with the hotel products and services and for this reason, some authors (Cantallops and Salvi, 2014) consider that customer online reviews mitigate the effects of information asymmetry and influence future customers' hotel booking intentions and behaviour.

3.2.2 Determinants of online hotel customer satisfaction: Hotel online textual reviews

Online customers' reviews (OCR) represent the customer voice to explore their experiences (Cheng and Jin, 2019; Zhang, 2019). Online customer reviews are used by hotels to understand customers' expectations and needs (Gu and Ye, 2014), to influence other customers´ decisions (Gao et al., 2018), to predict the overall customers satisfaction (Zhao et al., 2019) or to define the hotel preferences of customers (Li et al., 2015). Therefore, listening what customers say through online customer reviews is crucial to improve customer satisfaction. Considering the modern way of listening to consumers, the existing analytical methods have progressed. First by measuring word frequencies in small samples (Crotts et al., 2009) to later employing factor analysis in order to identify different factors based on word frequency (Xiang et al., 2015) to even generating more complicated machine learning tools such as sentiment analysis and topic modeling in order to extract deep meanings from large datasets (Ding et al., 2020). Moreover, some studies such as Li et al. (2013), who analysed hotel reviews using text mining technique and identified factors that influence customer satisfaction such as: accommodation, accessibility to transportation, closeness to tourist destinations, and cost effectiveness.

In addition, Dinçer & Alrawadieh (2017) study obtained results (referred to luxury hotels in Jordan) that were consistent with Berezina et al.'s categorization but also suggested further categories to their study, including the quality of the food and beverage, hotel policies, inadequate safety precautions, and misinformation. Hence, online reviews deliver the opportunity to precisely complete in-depth analyses of consumer behaviour based on normative online rating criteria (Nilashi et al., 2018).

However, despite hotel online reviews and ratings have become crucial for travellers' decision-making, a limited number of studies demonstrate the effectiveness of online ratings on consumers' decision-making and on travellers' preferences and satisfaction (Yu et al., 2017; Gavilan et al., 2018; Zhao et al., 2019). Consequently, very recently some researchers are becoming more and more interested in this stream of research (Metha et al., 2023; Song et al., 2022, Glavely et al., 2022; Hu et al., 2021, Phillips et al., 2020) as they consider that in the dynamic and competitive environment where hotels operate, examining online textual reviews is challenging because of the open structure of the reviews and the large amount of information that may be collected in them.

In this line, some authors supported that the technical attributes of the online textual reviews (for instance, the linguistic style, the subjectivity, diversity or length) can explain significant variations in hotel customer' ratings (Geetha et al., 2017; Zhao et al., 2019; Putranto et al., 2021). Related to this, previous studies have found inconsistency in hotel customers' opinions mined from their online textual reviews (OTR) and their ratings (Zhang et al., 2016). Some of them have found that the sentiments of OTR and customer ratings are highly correlated (Geetha et al., 2017; He et al., 2017; Antonio et al., 2017).

However, few studies explore the linguistic style or other technical attributes of the OTR in hotels (Zhao et al., 2019; Geetha et al., 2017) or in airbnb accommodation (Ding et al., 2020) and consequently, the relationship among other technical attributes of OTR like subjectivity, diversity, length and customer ratings is still under-explored (Zhao et al., 2019, Ding et al., 2020; Putranto et al., 2021).

To fill this gap in the literature, Zhao et al., (2019) examined online reviews from customers that stayed in city hotels in San Francisco (US) who posted their reviews in TripAdvisor and concluded that considering the limited resources, hotel managers should put more emphasis on the technical attributes implied by the online textual reviews that have: more subjective words expressing the customers' individual emotions (higher subjectivity), less number of diverse words (lower diversity), more advanced words (higher readability), more negative emotion (lower sentiment polarity) and longer length (contain more words). Specifically, the authors found that some technical attributes like subjectivity, readability and length influence in a significant and negative way the hotel customers ratings whereas other technical attributes like diversity and sentiment polarity influence customer ratings in a positive way. However, the context of this research included only OTR from urban hotels in San Francisco and the authors made a call for future research in other hotels in different contexts considering the fact that OTR can be influenced by the different languages and the cultural backgrounds of the hotel customers.

Table 3.1 Summary of studies on factors extracted from topic modeling

Study	Method	Factors extracted	Data sample
Kwon et al., (2021)	LDA	They found that the flight's "seat, service, and meal" are significant problems in the airline services.	More than 14,000 reviews from 27 airlines were collected from online reviews.
Putranto et al. (2021)	LDA	They observed that the most frequently discussed topics are services, price/food, facility, comfort, and location.	Over 50,000 samples were collected from 510 hotels in Indonesia
Ding et al. (2020)	STM	Internet connection, booking experience, transportation, sleep condition, group stay.	242,020 Airbnb reviews in Malaysia
Zhang (2019) Airbnb reviews	LDA	16 topics was extracted using LDA. Recommendation, clean room, late check-in, public transportation, parking and location, restaurant, and shopping.	They collected 2,799,420 reviews from 64,464 listings on the Airbnb platform in 10 American cities.
Dincer and Alrawadieh (2017) Luxury hotels	Content analysis	Food and beverage quality, hotel regulations, inadequate safety precautions, and misleading information.	Based on 424 negative TripAdvisor reviews for luxury hotels, content analysis was conducted.

Guo, Barnes, and Jia (2017)	LDA	In 5-star hotels; comfort, effective event management, and animal friendliness was the topics discussed while in 4 and 4.5-star hotels the topics are resort amenities, food quality, room size, and decoration.	big data set includes 266,544 online reviews for 25,670 hotels located in 16 countries
Berezina et al. (2016)	Text-mining	Business location (such as a hotel, restaurant, or club), room, furnishings, members, and sports.	2,510 hotel reviews for Sarasota, Florida, were collected from TripAdvisor
Li et al. (2013)	Text-mining	Accommodations, accessibility to transportation, closeness to tourist attractions, and cost effectiveness.	43668 online reviews for 774 star rated hotels
Gu and Ryan (2008)	Quantitative Survey	The quality of the beds, the bathrooms, the size and condition of the rooms, their location, their accessibility, the food and drinks, the auxiliary services, and the performance of the staff.	Total sample of 400 Beijing residents were interviewed

Since the tourism sector is a highly competitive industry, it would be very helpful for hotels to understand what customers are saying in online reviews in order to develop service improvement strategies. Therefore, the goal of this study is to effectively identify customer satisfaction and dissatisfaction and presenting marketing implications for the hotel industry to survive in the competitive tourism market. In this research, we used topic modeling and sentiment analysis to extract topics from related words in order to understand what luxury hotel guests are discussing online and how these topics relate to customer satisfaction.

3.2.3 Techniques for textual data analysis based on sentiment analysis and topic modeling

In recent years, some hospitality and tourism researchers are using complicated machine learning tools such as sentiment analysis and topic modelling in order to extract deep meanings from large datasets (Ding et al., 2020; Putranto et al., 2021; Song et al., 2022; Metha et al., 2023).

Given that online reviews and comments from customers have a significant impact on their purchasing behaviour, the need to use sentiment analysis and topic modelling approaches are rapidly expanding as a novel research technique to analyse in-depth opinions from consumers (Ban and Kim, 2019). Due to the great amount of online availability of data from hotel customers, recent research has actively used topic modelling to uncover the hidden meanings and topics from the text (Kwon et al., 2021) while other studies have used sentiment analysis which uses natural language processing (NLP) to better understand whether people have negative or positive opinion posted in reviews (Liu et al., 2012).

Topic modeling and sentiment analysis are the two main subcategories of text data analysis. While "sentiment analysis" refers to a set of techniques for identifying which emotions or feelings exist in the text, "topic modelling analysis" refers to a set of approaches for figuring out what the message hidden in the text is about (Blei al. 2003). The topic modelling approach allow extract topics from a corpus without needing a manual summarization and therefore reducing subjective biases. The topic modelling results provide not only topic-related keywords, but also, the proportional distributions of the different topics over each document. As a result, with this technique, it is possible to monitor changing patterns of customer perceptions to specific topics and it can improve the qualitative interpretation of textual data by providing information about the semantic relations of words from a corpus and the correlations of different topics (Ding et al., 2020).

In addition, Latent Dirichlet allocation (LDA) is used to analyse the large textual documents. The main premise is that each topic in documents is represented as a topic distribution represented by a distribution across the words. We choose the LDA model over other text analysis because of the following reasons: first, the LDA model succeeds at efficiently analysing massive amounts of data at a highly simplistic level, allowing us to discover the diversity of dimensions in online reviews. Secondly, by using the intensity of each dimension in the online reviews, LDA enables us to evaluate the pragmatic occurrence frequency of each dimension for instance when people use their own words to express their feelings or opinions about various hotel attributes, and therefore, LDA does not assume anything about the grammar or structural aspects of the text. (Guo et al., 2016 & Blei 2012). The underlying relationships between words and the context are disregarded by frequency-based approaches (Ahmad and Laroche, 2015).

Based on how frequently these topics are mentioned in online reviews, these topics are associated with their experiences and therefore consider as a crucial aspects of traveller satisfaction or dissatisfaction (Guo et al., 2017).

Apart from Topic Modelling, another important subcategory of text data analysis is Sentiment Analysis. Sentiment analysis has become very considerable in the service industry being polarity, valence, and arousal the most used metrics of sentiment (Mohammad, 2016). Customer sentiment in most situations relates to the feelings conveyed in text reviews by each customer (Geetha et al., 2017). Some authors found a high correlation between customers' overall rating and the sentiment polarity of their reviews (He, Tian, Tao, Zhang, Yan, and Akula, 2017; Geetha et al. 2017, Zhao et al., 2019). Geetha et al. (2017) concluded that there was a positive influence of customer sentiment polarity on customer ratings. However, the relationship between customer review sentiments and hotel ratings were not consistent across the different hotel categories.

Lee and Yu (2018) applied LDA to assess airport service quality with the analysis of more than 40.000 reviews from Google Maps and similarly, Palese and Usai (2018) applied LDA in the e-commerce industry to measure service quality. Guo, Barnes, and Jia (2017) used LDA to identify 19 topics that guests frequently discussed in hotel reviews, and they then looked at how these topics affected satisfaction scores. In 5-star hotels. comfort, effective event management, and animal friendliness were the topics more discussed while in 4 and 4.5-star hotels the topics were resort amenities, food quality, room size, and decoration.

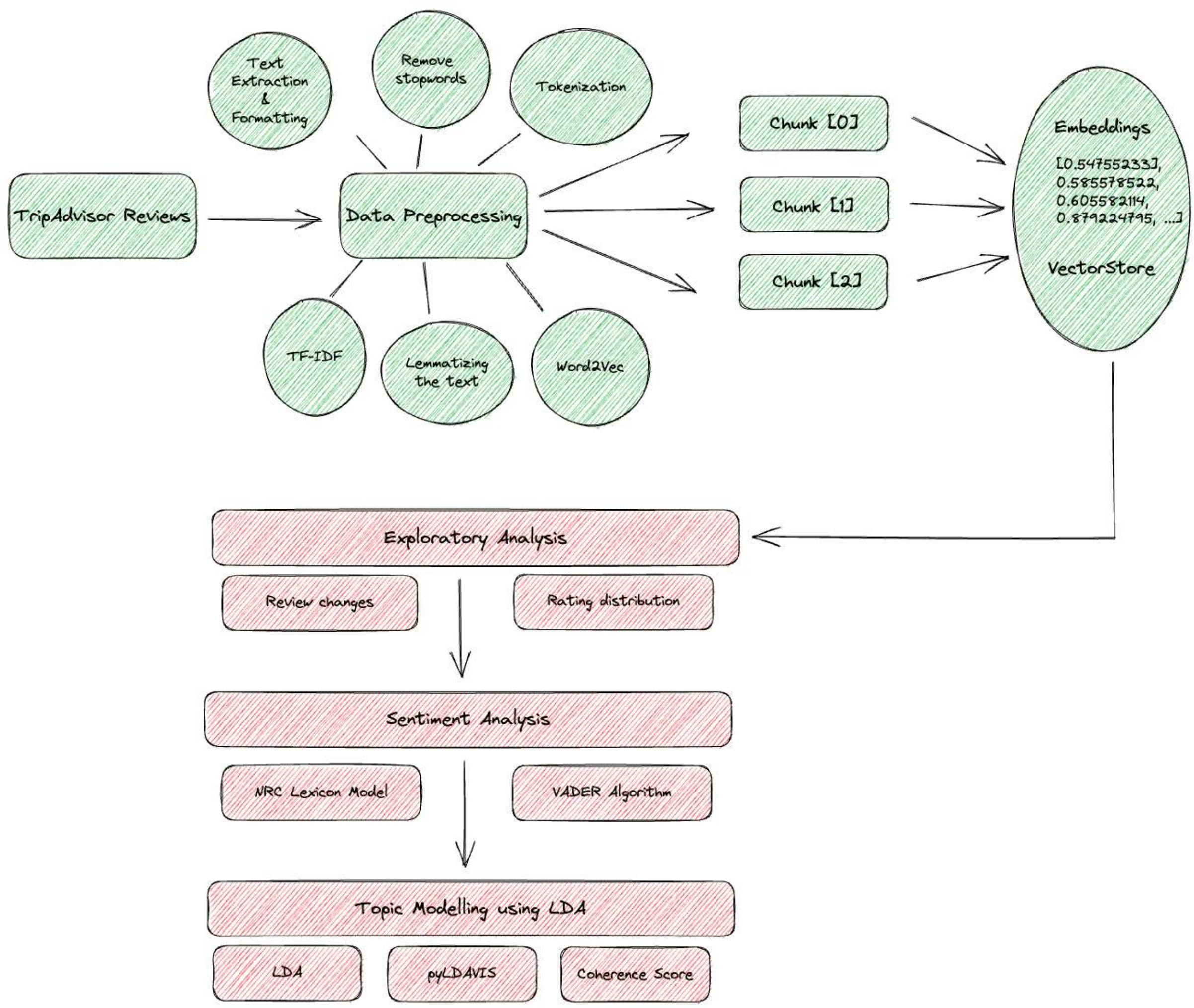
Also in the accommodation sector, recently, Zhang (2019) used topic modelling to extract 16 topics from Airbnb reviews and examined the impact on a property's listing performance. Although it has been confirmed that online review ratings have a significant impact on sales (Godes & Mayzlin, 2004), the effectiveness of topics derived from reviews of luxury hotel in predicting customer satisfaction has not been investigated in depth. Similarly, Zhao et al., (2019), highlight that the technical side of online textual reviews (namely the linguistic attributes) is still under-explored. They studied a big sample on San Francisco (USA) and recommend that future research extend their study by collecting more samples for multiple cities. Also, following Zhao et al.'s recommendation, very recently, Putranto et al. (2021) applied topic modelling and hotel ratings prediction based on customer reviews in Indonesia and they addressed the need of future studies to investigate factors that consider the changing perception of hotel users in order to contribute to predicting the variation of the customer preference and therefore be able to improve customers' satisfaction.

In an attempt to overcome most of these authors' recommendations, our main research objective is to provide an application of social media analytics in luxury hotels in the Canary Islands (Spain) using Topic Modelling and sentiment analysis during a long period before the pandemic COVID-19 arised (from 2002 to 2019).

Since the tourism sector is a highly competitive industry, it would be very helpful for hotels' managers to understand what customers are expressing in online reviews in order to develop service improvement strategies. Therefore, the goal of this study is to effectively identify customer satisfaction and dissatisfaction in the previous stage to the pandemic, through a longitudinal analysis and to present marketing implications for the hotel industry to survive in the competitive tourism market in the specific context of Spain.

In this line, our research fills a gap in the literature by exploring perceived service quality attributes in hotels and by predicting the variation of customer preferences in luxury hotel users in Spain. In this study, we will investigate the performance of LDA to predict customer satisfaction ratings and the quality of extracted service quality attributes.

Figure 3.1: The framework of the data analysis



3.3 Methodology

Customers can use online reviews, which are a common type of eWOM, to share information about their opinions of products and services and to search up reviews of the services they are interested in. As a result, TripAdvisor was selected for this study as the most widely used source of hotel ratings. Travelers share their experiences on TripAdvisor, which is regarded as the largest travel website in the world.

Consumers use these reviews to make a decision and help them to plan their own trips. TripAdvisor now has over 860 million reviews and opinions, which cover about 8.7 million hotels, airlines, activities, and restaurants. With an average of 463 million visitors each month[2], TripAdvisor is also the largest travel platform in the world.

In "The Travel & Tourism Competitiveness Report", 141 countries were analysed, and Spain came out on the top 3.[3] One of the most popular tourist destinations in Spain is the Canary Islands. Over 14 million tourists were hosted there in 2019. (Gran Canaria Tourism board, 2021). Due to their popularity with German, British, and Scandinavian tourists, the Islands are well-known throughout northern Europe. This study focuses on Canary Islands hotels through TripAdvisor as they are among the most popular tourist destinations in Europe (Hannonen et al., 2023; Aguiar-Quintana et al., 2022; Ahani et al., 2021; Estay-Ossandon et al., 2018). Accordingly, the major nationalities reported in the FRONTUR survey (2019) amount for more than 90% of the total international visitors in the Canary Islands.

[2] Source: Tripadvisor internal log files, average monthly unique visitors, Q3 2019
[3] https://www.weforum.org/reports/travel-and-tourism-development-index-2021/digest

Overall, Germany, the United Kingdom and the Scandinavian country are the major inbound markets. In the case of Gran Canaria Island, it receives an average of 3 millions of tourists per year which represents 300% its local population and tourism contribute. With 15,000 beds spread across 20 hotels in Gran Canaria, Fuerteventura, Germany, Austria, and the Dominican Republic under its two chains, Currently, Lopesan is among the top 10 hospitality businesses in Spain and the Canary Islands: Lopesan Hotels & Resorts and IFA Hotels.[4]

3.3.1 Data Collection and Pre-processing

This study took into account three luxury hotels: the Lopesan Costa Meloneras (Costa Meloneras), Lopesan Baobab (Baobab), and Lopesan Villa del Conde Hotel (Villa del Conde). The range of four to five stars given to the selected hotels demonstrates that the sample represents a homogeneous group of hotels in terms of their service ratings, which is essential to confirm the data accuracy.

Web scraping extracted 8,376 reviews of Lopesan chain hotels in Gran Canaria. Beautiful soup library in Python was used to extract the customer reviews. The reviews extracted are from October 2022 to February 2019. Table 3.2 displays the total number of reviews collected for each hotel as well as the rating mean. In Table 3.3, the rating frequencies are demonstrated.

The next step was data preprocessing, which involves data cleaning, where the unnecessary clutter gets removed from the data set. For example, all the stop words and special characters are removed. Tokenization, stemming, and lemmatization are a part of the preprocessing stage. Tokenization involves tokenizing data into terms.

[4] https://www.lopesan.com/en/corporate/lopesan-history/(30/11/2022)

The word variations were reduced by using stemming and lemmatization that converted the inflected words to a common base.

Table 3.2 The review numbers and mean ratings for each hotel

	Lopesan Baobab	Lopesan Costa Meloneras	Lopesan Villa del Conde
Review numbers	3967	2648	1773
Rating Mean (SD)	4.50 (0.84)	4.39 (0.94)	4.49 (0.85)

3.3.2 Descriptive statistics

The collected data was analysed using Python software. The likelihood that latent topics are present throughout the entire document was calculated using statistical text processing approach. The topic modelling LDA approach (Latent Dirichlet Allocation) has been used in this study. In addition, sentiment analysis was used in this paper to investigate the positive and negative words in consumer reviews. There are 3 stages that have been used for the topic modelling; first, we collected the data needed for the analysis. Second, we clean the data to be used for the topic modelling and finally, we conducted the analysis.

Figure 3.2 shows how the number of reviews has changed over time for each hotel/year.
As we can see from figure 3.2, the number of reviews since 2012 has been increased. Although Baobab was the most recently constructed hotel, it receives the highest number of reviews, which indicates that hotel guests have a seasonal pattern of behaviour.

Figure 3.2 Review numbers of each hotel

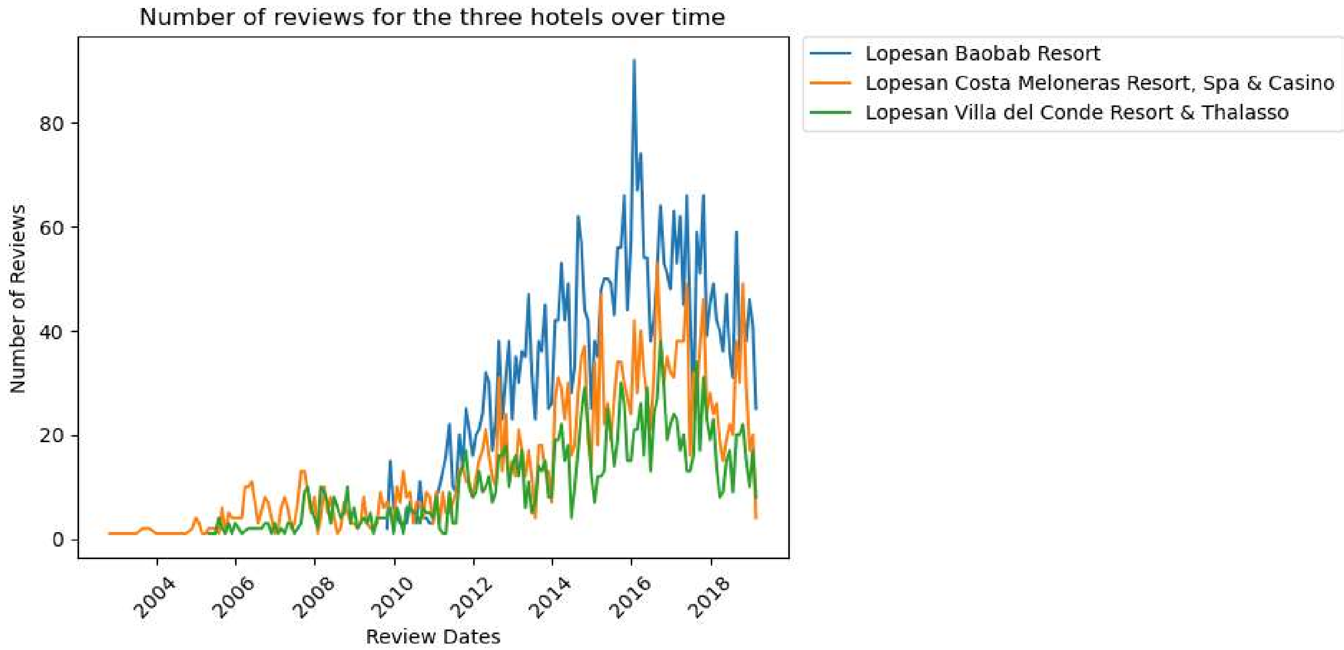
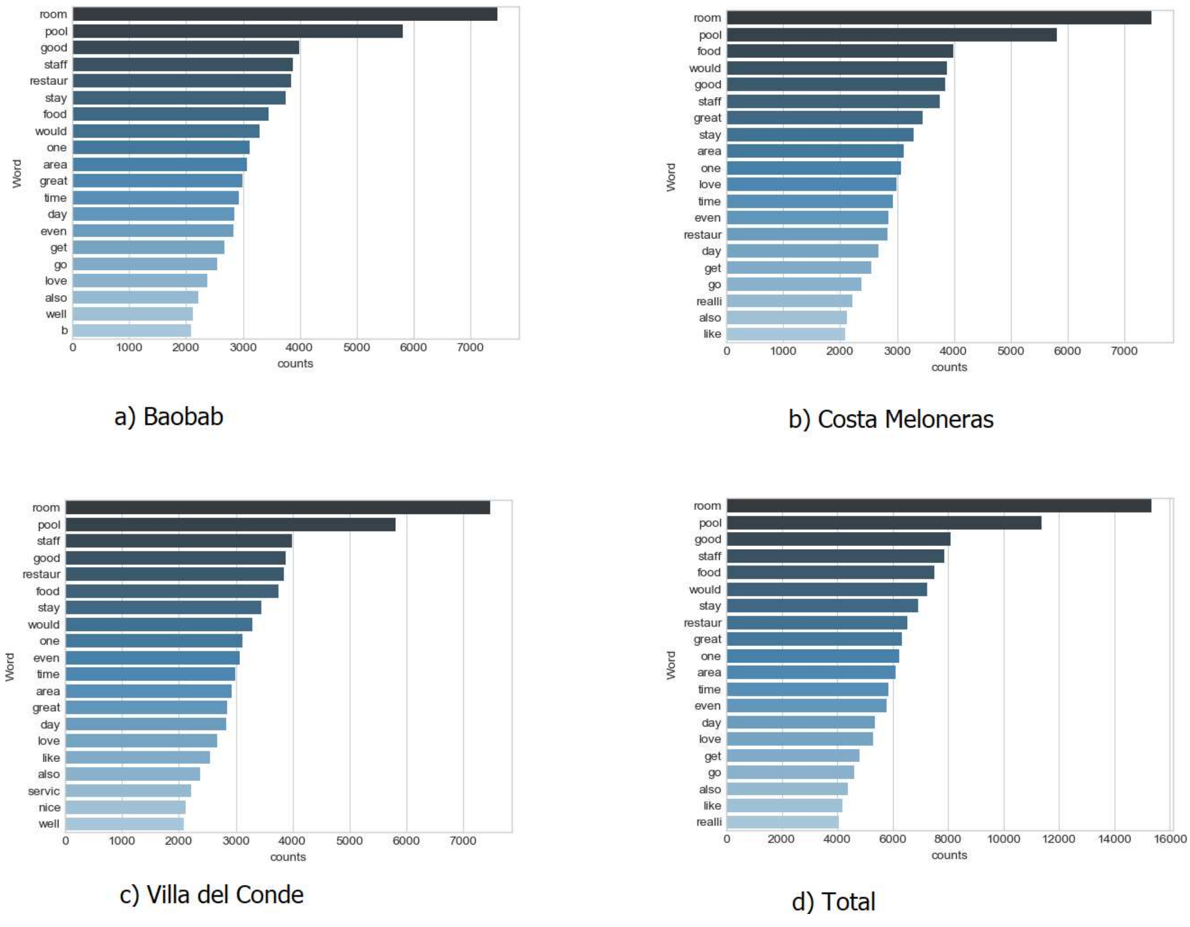


Figure 3.3 displays the 20 words that appear the most frequently across all reviews from the three hotels as well as the total reviews. A Bag of Words (BoW) model has been used in this study. With this approach, each word count is considered a feature as we analyse the histogram of the text's words. [For more information, Zhang et al., 2010 and Wu et al., 2010]. In order to generate the histogram, we clean the data by removing the unnecessary words (such as "a," "is,") and reduces word variations that are related to inflection and derivation to a base form (stemming and lemmatization).

As we can see, some characteristics vary in importance depending on the hotel, but remain consistent across the chain as a whole.

Figure 3.3 Words that appear most commonly in reviews



a) Baobab

b) Costa Meloneras

c) Villa del Conde

d) Total

Table 3.3 Frequency of ratings by hotel

Frequencies	**Rating**				
	1	2	3	4	5
Baobab	61	81	290	902	2633
Costa Meloneras	67	80	206	683	1612
Villa del Conde	28	47	116	411	1172

3.3.3 Sentiment Analysis

Sentiment analysis is the process of identifying and classifying the thoughts presented in a review in order to identify whether the customer has positive, negative, or neutral attitudes and emotions. Sentiment analysis can be used to help hotel managers to better understand the reviews' customer sentiment and be able to find and extract information from any context by using the text-mining technique (Hogenboom, Heerschop, Frasincar, Kaymak, & de Jong, 2014). The growing popularity of online review platforms among consumers has given sentiment analysis more power. Additionally, compared to surveys and opinion polls, it has the advantage of being more time effective (Geetha et al., 2017).

For the purpose of this study, a value range of -1 (extremely negative) to 1 (extremely positive) was used to calculate the sentiment polarity of each review. This feature measures the sentiment intensity using a lexicon rule-based model for sentiment analysis called VADER (Valence Aware Dictionary and sEntiment Reasoner) (Hutto & Gilbert, 2014). In Figure 3.2, the percentages of positive (polarity > 1), neutral (-1 < polarity < 1), and negative (polarity < -1) reviews for each hotel are presented.

Figure 3.4 Proportions of positive, neutral, and negative scores

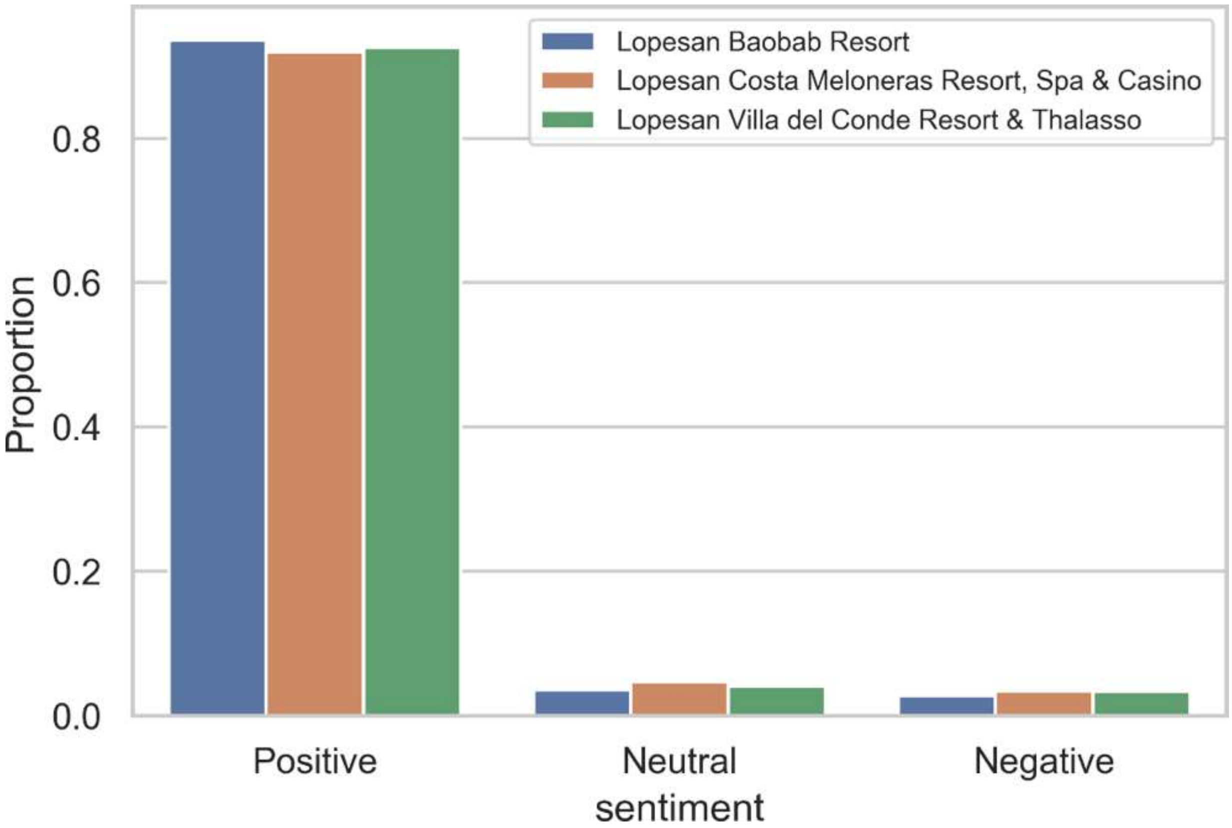


Figure 3.4 displays the percentage of reviews' words that are attached to each sentiment. As we can see that the majority are positive sentiments which means that the satisfied customers with the hotels are more than dissatisfied ones.

Figure 3.5 is used to show the data distribution using a five-number summary for each emotion. The first quartile (Q1) minus 1.5 times the difference between the second (Q2) and third quantiles (Q3) is used to recognise the bottom of the whisker. Interquartile Range (IQR) refers to this variation. The Q2 is represented by the bottom of the box. The box's dividing line is the median, third quartile (Q3) is at the top of the box, and Q3 + 1.5*IQR is at the top of the whisker. The distribution's outliers are represented by the dots outside the whiskers.

The main feelings expressed in the reviews of the three hotels are Joy, anticipation, and trust.

Figure 3.5 Boxplot of each emotion for each hotel.

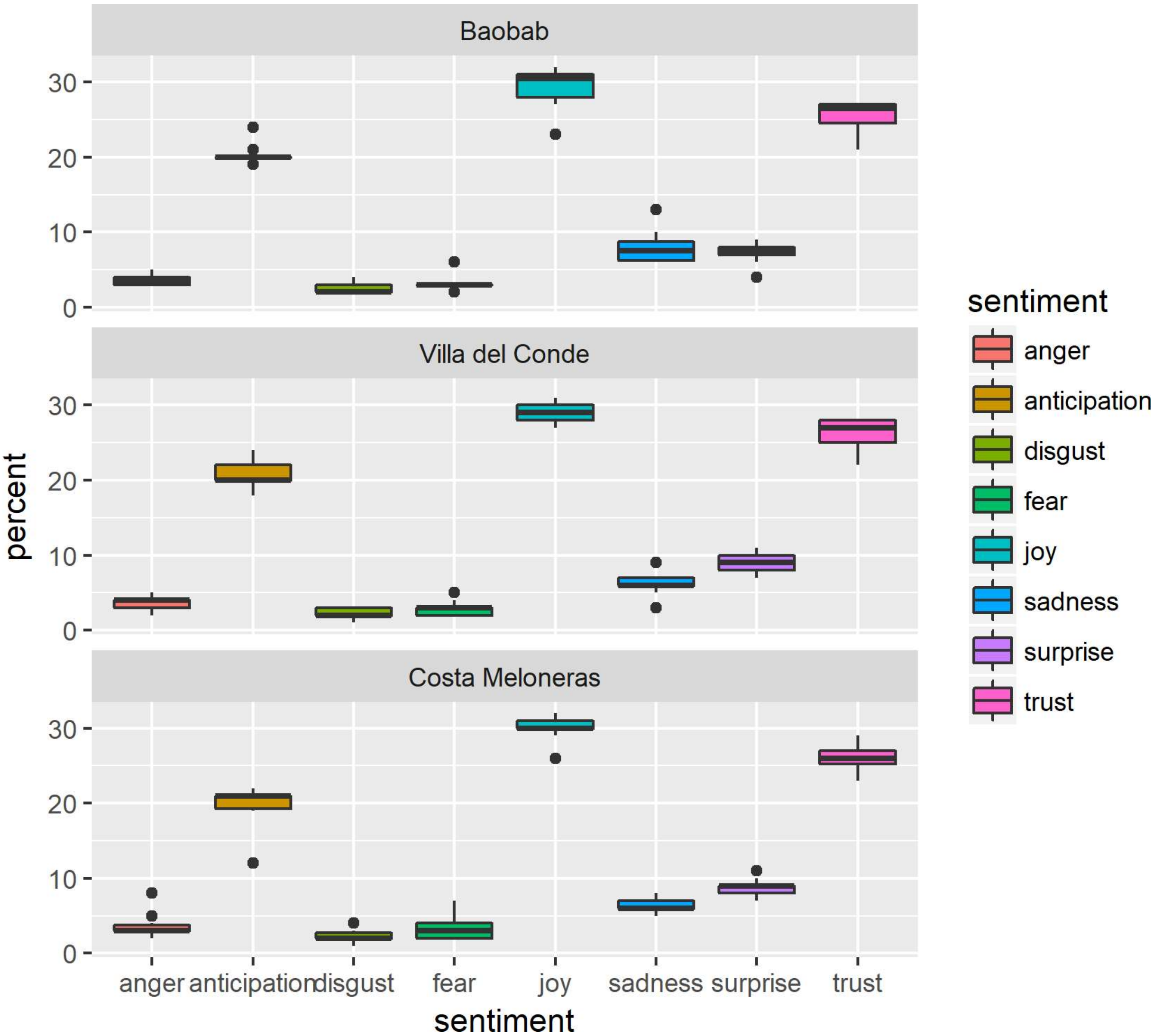


Figure 3.6 Emotions words usage over time

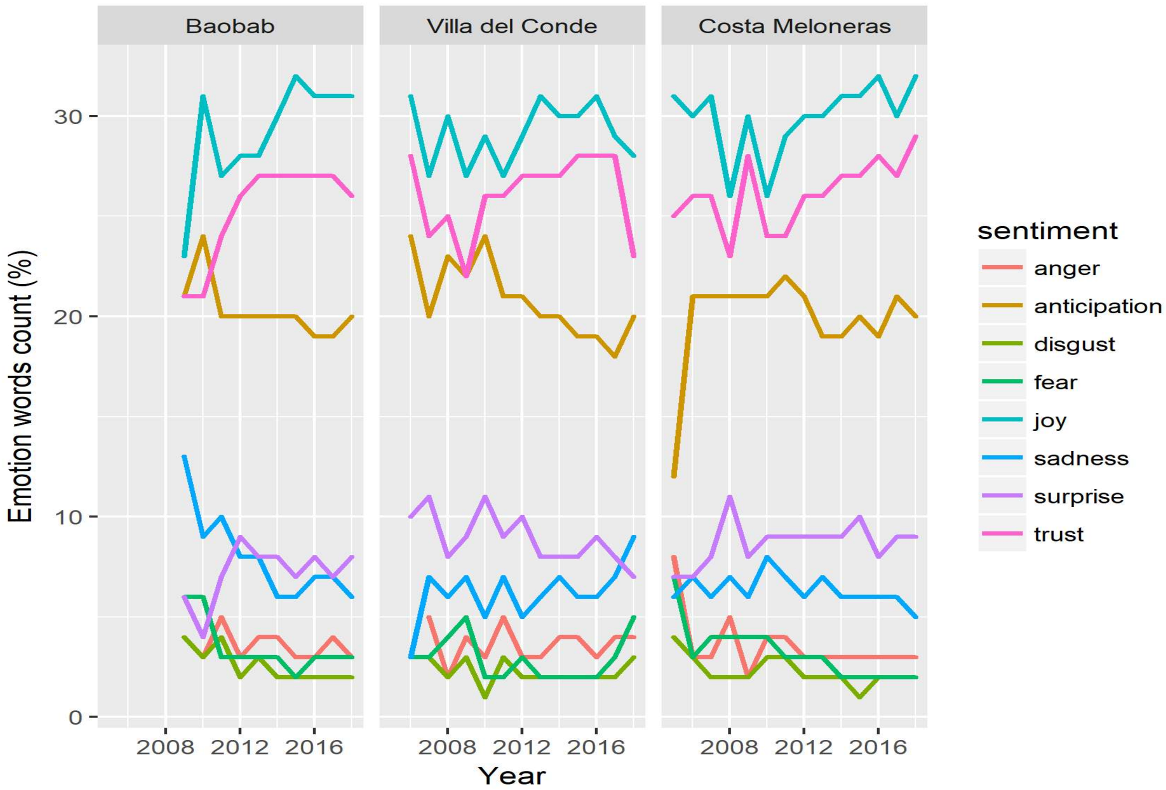


Figure 3.7 Distribution of polarity and subjectivity

We can see the distribution of polarity and subjectivity in Figure 3.7 where the x-axis shows the polarity score ranges from -1 (more negative) to +1 (more positive) and the y-axis shows the subjectivity score and ranges from 0 (objective in expressing facts and information) to 1 (subjective in expressing opinion and emotions). From figure 3.7, we can observe that most of the reviews are positive with a concentration of reviews around a polarity score of 0.5 and a subjectivity score of 0.5. This suggests that the majority of hotel reviews are generally positive and somewhat subjective. However, we can also observe some reviews with negative polarity scores, indicating dissatisfaction with the hotel, and some reviews with high subjectivity scores, indicating strong personal opinions and emotions in the review.

3.3.4 Topic Modelling using LDA

Topic modelling is a machine learning technique that use the identification of topics within a collection of documents (Blei et al., 2003). This method sheds light on the narratives that are present in the data. In this study, the text was categorized into categories using Latent Dirichlet Allocation (LDA), and the frequency of words used by various reviewers in their hotel ratings was looked at. Lei et al. (2003) describe LDA as a potent method for separating latent semantic structures from text input and identifying underlying constructs using Bayesian learning. The algorithm has two steps: the first assigns pertinent themes to each text, and the second gives each topic's phrases a probability distribution (Blei et al., 2003).

The LDA model is enacted on the TripAdvisor dataset of 8,376 customer reviews. For calculations, we passed 100 samples with 20 topics, through 100 iterations. The model generates the topic distribution. Table 3.4 presents the sample topics and word distribution. These topics have keywords of ten words. All the 20 dominant topics are identified by adopting the topic modeling technique. LDA provides relevant topics having the maximum weights. The topics are presented by grouping the keywords with the maximum possible occurrence, distinguishing all topics from one another.

3.3.5 The Logistic Regression (STATA)

Due to the fact that our dependent variable is categorical, we used the Logistic ordinal regression. This regression is a member of the family of generalised linear models and can be applied when the outcome variable is categorical (McCullagh & Nelder, 1989).

To understand how predictors affect customer satisfaction scores, ordinal logistic regression was used. Review polarity and length, emotions embedded in review text, prior average rating, and topic dummies are the variables that we considered in this study. The models were run both ways and then compared. Online individual rating is our dependent variable that we used to test the model. It shows the overall customer rating for the hotel. Ratings are range from 1 (low) to 5 (high). Customer satisfaction with the product is measure by the online ratings (Engler et al., 2015).

3.3.6 Variables

The reviewer's rating (Y_{ijt}) on a scale of one (terrible) to five (excellent) stars was the dependent variable (Li et al., 2019, 2020).

Independent variables

Social influence was operationalized as the average of prior ratings of the hotel before the focal reviewer (Sridhar & Srinivasan, 2012, pg. 75). Thus, the independent variable "social influence" was measured as the prior average rating ($PriorAvgRating_{ij(t-1)}$) and it could have values between 1 and 5. Since TripAdvisor shows visitors the average prior ratings rounded to the nearest 0.5 points (i.e. bubble rating). Prior Volume of Reviews was also generated by calculating the sum of review count for each hotel.

Topics are the dummy variable, and we have identified 20 topics based on the coherence score. Review texts were used to extract characteristics linked to consumers' product experience. The first review content-specific moderators are anger and joy, which were measured as the total number of emotional words associated with anger and joy in EmoLex. Hotel names were also measured as a nominal variable, indicating 1 (Costa Moleneras was the reference one), 2 for Baobab, and 3 for Villa del Conde.

Control variables

We controlled for review length, which was assessed by the logarithm of the number of words in the review (Poncheri et al., 2008; Li et al., 2020). The natural logarithm transformation was used to normalize this measurement because the range of number of words was highly scattered.

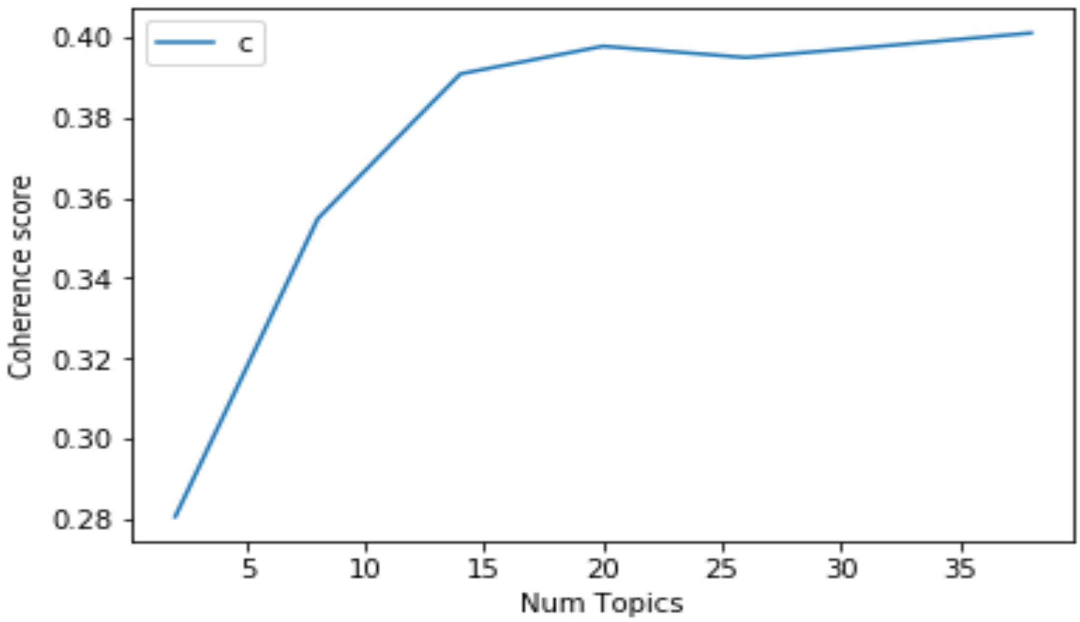
The *polarity* score is ranges from -1 (most extreme negative) to 1 (most extreme positive), higher score implies that the text has a more positive sentiment. A score of 0 indicates that the sentiment is neutral (Geetha et al., 2017).

3.4 Results

3.4.1 Topic modelling

Topic modelling is a text mining technique uses machine learning to extract themes from text documents (Blei, 2012). Table 3.4 shows the top 20 topics for the three hotels using LDA algorithm (latent Dirichlet allocation). Coherence and perplexity statistics are used to determine the ideal number of topics (figure 3.8). LDA is the most widely used algorithm for studying a large collection of documents. The fundamental concept is that each topic in a document is represented as a topic distribution, with a word distribution to describe it (Kwon et al., 2021).

Figure 3.8 Perplexity and Coherence score graph



According to the plot above, coherence score rises with the number of topics, declining between 20 and 25. Therefore, we choose 20 topics as the optimal number of topics as going for higher number of topics may contain repeated keywords.

The name of the topic was determined by logically connecting the most frequently used words for a topic; therefore, we were able to determine the name of a topic using the set of keywords that the LDA suggested. For instance, the topic #3 name "hotel size and design" is based on the words "Area, large, pool, design, huge" all of which appear at the top of the list.

Table 3.4 Topics and keywords

Topic # and label	Relevant Keywords
1: View and romantic surroundings	Room, pool, view, area, restaurant, hotel, sea, main, large, quiet
2: Service quality	Good, excellent, service, quality, spa, high, facility, standard, restaurant, location
3: Hotel size and design	Area, large, resort, pool, garden, style, feel, huge, design, comfortable
4: Checking service	Room, reception, check, day, arrive, leave, book, give, service, wait
5: Convenience of location	Walk, beach, restaurant, shop, bar, hotel, minute, close, front, plenty
6: Returning customers	Hotel, visit, stay, year, time, return, holiday, staff, fantastic, back

7: Family with kids	Child, family, pool, kid, great, couple, adult, love, young, time
8: Hotel facilities (especially pool)	Good, nice, pool, great, hotel, lot, big, clean, place, food
9: Staff quality/attitude	Staff, friendly, clean, food, helpful, excellent, choice, ground, plenty, return
10: Restaurant quality	Restaurant, food, hotel, dinner, evening, good, pool, eat, week, cold
11: Hotel lovers	Great, amazing, fantastic, staff, lovely, food, beautiful, back, perfect, love
12: Hotel fers	People, time, review, <u>bad</u>, issue, point, <u>problem</u>, thing, read, <u>complain</u>
13: Pay problems	Room, pay, hotel, free, <u>charge</u>, safe, <u>extra</u>, drink, price, day
14: Satisfied hotel guests	Hotel, stay, lovely, week, night, find, back, time, feel, eat
15: Nightlife and Hotel entertainment	Entertainment, evening, night, good, bar, show, hotel, German, watch, bit
16: Hotel amenities	Hotel, pool, bed, find, towel, sunbed, area, day, people, plenty
17: Breakfast and food facilities	Breakfast, restaurant, night, bar, food, dinner, drink, meat, eat, selection
18: Room services	Room, day, bed, water, change, double, end, towel, bottle, clean

19: Problem issues	Room, shower, open, reception, floor, bathroom, bath, light, door, noise
20: Crowdedness or hotel guest quality	Hotel, guest, make, star, feel, experience, place, staff, stay, number

Figure 3.9 Topic Comparison Heatmap for Three Hotels

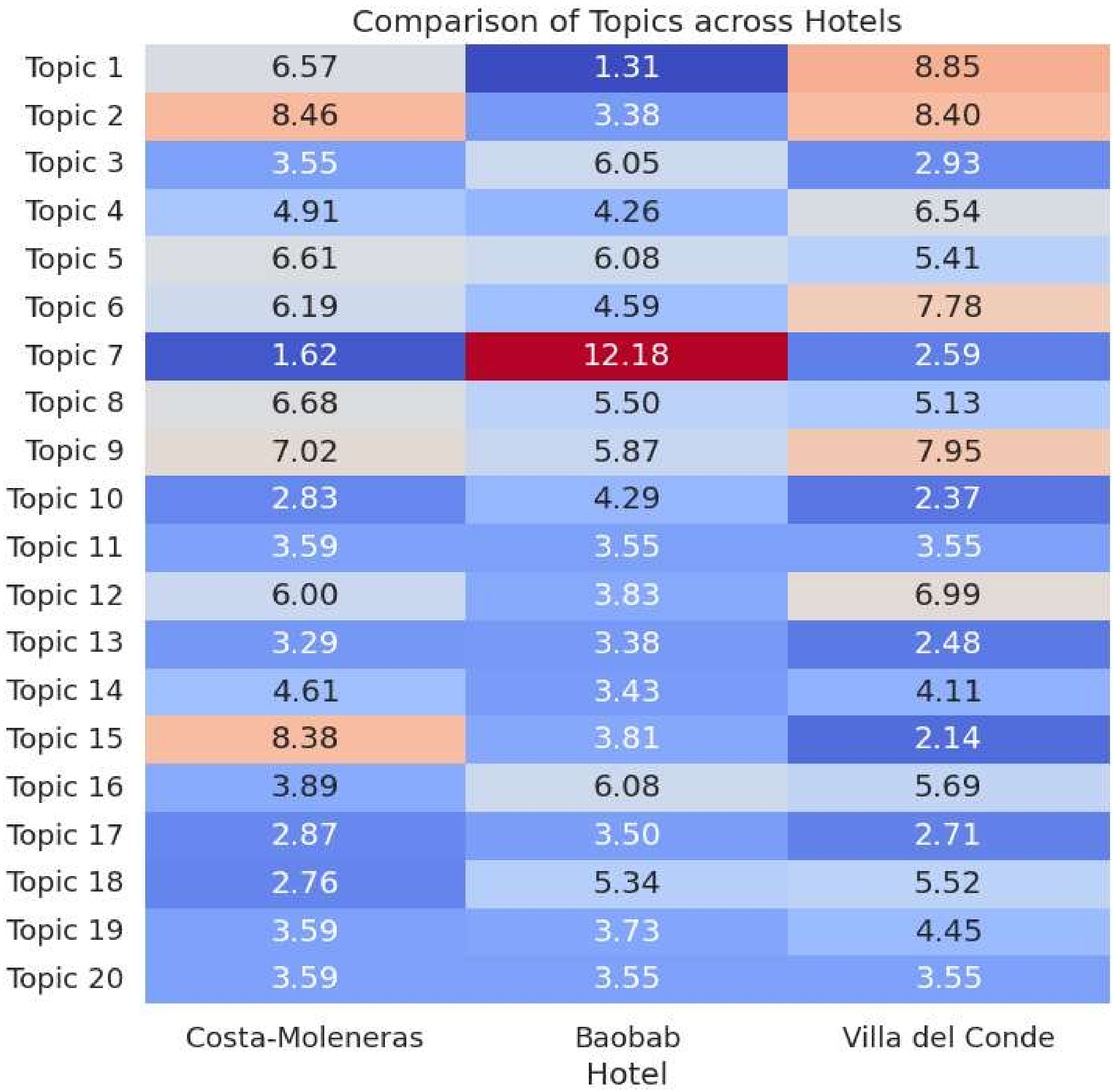
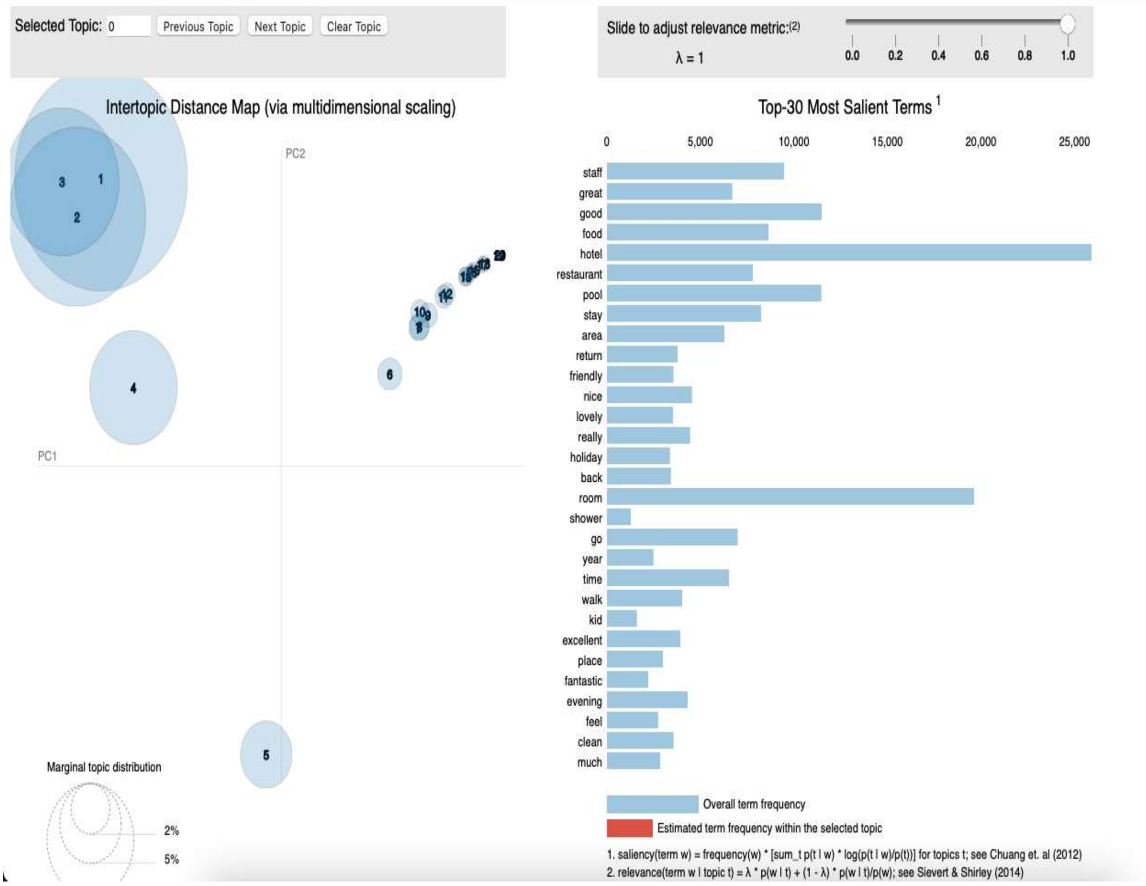


Figure 3.9 shows the percentage of each topic in each hotel. Heatmaps were utilized as shown in Figure 3.9 to compare the topic's distribution between the three hotels. When displaying data visually, heatmaps use color-coded cells to represent the various values in a matrix. Darker colors suggest greater correlations between two variables, while the colors in the matrix cells itself represent the weight or intensity of the association between two variables.

A darker color on the heatmap shows greater significance values, therefore suggesting a greater connection between a given topic and a particular hotel. One can notice that the hotel "Baobab" column has the most weight (dark blue) when we look at row 2 (Topic 2). According to this, Topic 2 is more closely connected to the Baobab hotel than it is to the other two hotels. Likewise, row 7 (Topic 7) demonstrates that the "Baobab" hotel column once again holds the highest weight. Therefore, Topic 7 shows a greater and highest connection to the Baobab hotel than with other hotels.

Figure 3.10 LDA topic modelling visualisation



As shown in Figure 3.10, the LDA model was visualized using the output from the pyLDAVis package in the form of an intertopic distance map, which helps to view the topics and keywords. The topics are represented by the multiple bubbles on the map. Relevant topics with the highest weights are provided by LDA. The terms with the highest frequency are grouped to present the topics, setting each one apart from the others. The size of the bubbles affects how prominent the topics are. Additionally, the overlap of the bubbles shows that the overlapping topics that share certain similar terms. The LDA visualization's bar chart that is included with the map shows the local frequency of keywords that are included in topics.

The term frequency bar charts are located on the right side of the visual, and the global topic view is located on the left. Topics are well separated if they do not overlap with one another. Topic 1,2,3,4,5 are the most discussed topics therefore, it would be interesting to check whether these topics have influence on satisfaction ratings or not.

Table 3.5 Results of multilevel ordered logit regression (standardized values used)

	M1	M2	M3	M4
Anger	-0.557***	-0.470***	-0.243***	-0.260***
Joy	0.358***	0.316***	0.641***	0.642***
Prior Average Rating	0.118***	0.114***	0.0736**	0.0986***
Prior Volume Reviews	0.0644**	0.0513*	-0.265***	-0.251***
Review Length	-0.290***	-0.154**	-0.123*	-0.130*
Polarity	0.644***	0.547***	0.522***	0.520***
Baobab		0.661***	0.342**	0.328***
VillaDelConde		0.404***	0.302***	0.286***
Topic 1		0.705***	0.714***	0.705***
Topic 2		1.160***	1.200***	1.190***
Topic 3		0.897***	0.884***	0.879***
Topic 4		-0.372*	-0.272	-0.263
Topic 5		1.320***	1.277***	1.265***
Topic 6		1.630***	1.651***	1.639***
Topic 7		0.725***	0.640***	0.636***
Topic 8		0.783***	0.758***	0.751***

Topic 9		1.766***	1.750***	1.736***
Topic 10		-0.245	-0.265	-0.266
Topic 11		2.402***	2.387***	2.379***
Topic 12		0.126	0.254	0.264
Topic 13		-0.326*	-0.364**	-0.373**
Topic 14		0.860***	0.930***	0.918***
Topic 15		0.437**	0.465**	0.454**
Topic 16		0.607***	0.600***	0.600***
Topic 17		0.821***	0.742***	0.734***
Topic 18		-0.342*	-0.301	-0.301
Topic 19		-0.0899	-0.0324	-0.0299
Anticipation			0.285***	0.281***
Disgust			-0.280***	-0.285***
Fear			-0.0170	-0.0112
Sadness			-0.140**	-0.143**
Surprise			0.0768	0.0746
Trust			0.134*	0.131
PriorAvgRating*anger			-0.0372	-0.0342
PriorAvgRating*joy			-0.0473	-0.0491
PriorVolumeReviews*anger				0.0705*
PriorVolumeReviews*joy				-0.0520
cut 1	-4.731***	-4.205***	-3.950***	-3.969***
cut 2	-3.689***	-3.157***	-2.882***	-2.901***
cut 3	-2.341***	-1.779***	-1.472***	-1.491***

cut 4	-0.596***	0.0822	0.425***	0.409**
N	8386	8386	8386	8386

Findings from the ordinal regression model are summarized (in Table 3.5).

Model 1 is the model with only control variables. Review length was significantly related to satisfaction ratings in a negative coefficient (b=-0.290) which means that when consumers feel dissatisfied, they tend to write longer reviews to express their negative emotions while sentiment polarity was significantly positive related to the ratings (b=0.644).

Model 2 used to add the 20 topics identified by the LDA topic modelling in order to examine the effect of each hotel on the rating satisfaction. Model 3 & 4 investigated the interaction effect between the prior average rating and two emotions (joy & anger) and the interaction effect between prior volume reviews and emotions.

Model 3 shows a negative interaction between prior average rating and both emotions (b=-0.0372; b=-0.0473) mean that the effect of emotions on customer satisfaction is smaller when the prior average rating is higher. In other words, customers who previously had a good experience with the hotel may be less affected by sentiments and emotions during their current stay. In contrast, customers who have had a bad experience in the past might be more likely to experience an emotional impact during their current visit. Model 4 shows a positive interaction between prior volume reviews and anger emotion (b=0.0705) while a negative interaction between prior volume reviews and joy emotion. This indicates that the impact of anger emotion on customer satisfaction increases with the number of previous reviews (customers are more likely to be influenced by negative feelings stated in online evaluations when there are plenty of prior reviews).

Contrarily, when there are more previous reviews, the impact of the joy emotion on customer satisfaction is smaller (i.e., customers are less likely to be influenced by favorable emotions stated in other customers' reviews when there are numerous prior reviews).

Hotel dummies had also important role as more people were satisfied with the Baobab hotel services (b = 0.321) more than with Villa del Conde and Costa Meloneras. Topic 4 has a negative effect (b = -0.366), thus managers should take actions regarding check-in service to better have smooth operation.

Table 3.6 shows the results of the multiple regression model to investigate the impact of different topics on the satisfaction ratings where the DV is the hotel categories where 1=Costa Meloneras hotel (REFERENCE), 2 = Baobab hotel, and 3 = Villa de Conde hotel

To summarize the results of Table 3.6:

Topic 1 (View and romantic surroundings)

- Baobab: Customers who mentioned this topic have significantly lower satisfaction (coefficient = -1.539) compared to Costa Moleneras.
- VillaConde: Customers who mentioned this topic have slightly higher satisfaction (coefficient = 0.273) compared to Costa Moleneras, but it is not statistically significant.

Topic 2 (Service quality)

- Baobab: Customers who mentioned this topic have significantly lower satisfaction (coefficient = -0.894) compared to Costa Moleneras.

Topic 3 (hotel size and design) theme was discussed more for Baobab and less in Villa del Conde compared to Costa Moleneras.

Topic 6 (Returning customers)

- Baobab: Customers who mentioned this topic have significantly lower satisfaction (coefficient = -0.360) compared to Costa Moleneras.

Topic 7 (Family with kids)

- Baobab: Customers who mentioned this topic have significantly higher satisfaction (coefficient = 2.056) compared to Costa Moleneras.

Topic 11 (Hotel lovers)

- VillaConde: Customers who mentioned this topic have significantly lower satisfaction (coefficient = -0.698) compared to Costa Moleneras.

Topic 13 (Pay problems)

- Baobab: Customers who mentioned this topic have significantly lower satisfaction (coefficient = -0.508) compared to Costa Moleneras.

Topic 16 (Hotel amenities)

- Baobab: Customers who mentioned this topic have significantly lower satisfaction (coefficient = -0.848) compared to Costa Moleneras.
- VillaConde: Customers who mentioned this topic have significantly lower satisfaction (coefficient = -1.534) compared to Costa Moleneras.

Topic 19 (Problem issues)

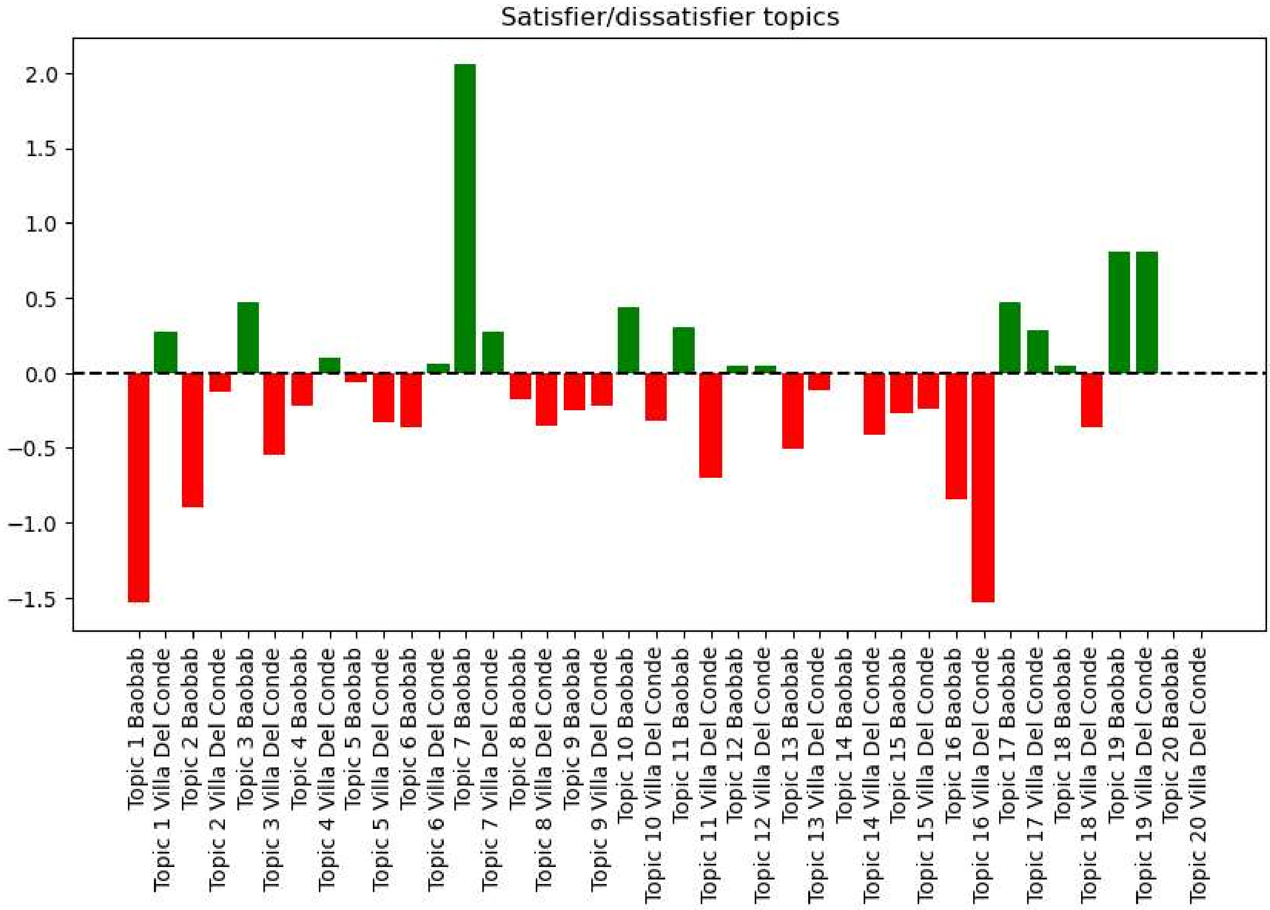
- Baobab: Customers who mentioned this topic have significantly higher satisfaction (coefficient = 0.809) compared to Costa Moleneras.
- VillaConde: Customers who mentioned this topic have significantly higher satisfaction (coefficient = 0.810) compared to Costa Moleneras.

Table 3.6 Satisfier/dissatisfier topics for each hotel (DV=Hotel (1: Costa Moleneras (REFERENCE), 2: Baobab, 3: Villa Del Conde)

	Baobab	**VillaConde**
Topic 1	-1.539***	.273
Topic 2	-.894***	-.129
Topic 3	.466*	-.545*
Topic 4	-.219	.094
Topic 5	-.063	-.328
Topic 6	-.360*	.058
Topic 7	2.056**	.271
Topic 8	-.183	-.358
Topic 9	-.246	-.220
Topic 10	.433*	-.326
Topic 11	.301	-.698**
Topic 12	.045	.049
Topic 13	-.508**	-.115

Topic 14	-.001	-.420
Topic 15	-.272	-.238
Topic 16	-.848***	-1.534***
Topic 17	.465*	.282
Topic 18	.052	-.364
Topic 19	.809***	.810***
Topic 20	Ref.	Ref.
Prior avg. rating	.691***	1.541***
Polarity	.138***	.048
anger	.017	.005
disgust	-.008	-.036
Fear	.112*	-.107
sadness	.259***	.234 ***
joy	.036	-.072
anticipation	.151	.093
surprise	-.263***	-.043
trust	-.169*	.297**
Review length	.033	-.194*
constant	.513***	-.377*

Figure 3.11 Satisfiers/dissatisfier topics



Another way of visualizing the results from Table 3.7 can be shown in Figure 3.11 where:

- Topic 7 (Family with kids) is the most frequently discussed themes in luxury hotel reviews.
- Topic 17 (breakfast/food facilities), topic 13 (pay problems), topic10 (food/restaurant quality) are less frequently discussed themes in luxury hotels.

- Topic 4 (Checking service), topic 13 (pay problems), and topic 18 (room services) are considering as dissatisfiers topics where topics 10 (Food/Restaurant quality), topic 12 (hotel complainers) and topic 19 (Problem issues) did not influence satisfaction ratings compared to topic 20 (Crowdedness/hotel guest quality).
- Topic 11 (hotel lovers), topic 9 (staff attitude/quality), topic 6 (returning customers), topic 5 (convenience of location) are the most important delighted customers topics.
- Topic 16 (hotel amenities) discussed less for Baobab and Villa del Conde compared to Costa Moleneras.
- Topic 19 (problem issues) discussed more for Baobab and Villa del Conde compared to Costa Moleneras.

3.5 Discussion

The objective of this study is to identify the factors that influence customer satisfaction using topic modeling and sentiment analysis of big data. We used an exploratory approach to investigate the elements that affect the customer's satisfaction by examining online reviews written by customers who have visited the three luxury hotels.

The results of Model 1 indicate that both review length and sentiment polarity significantly impact customer satisfaction ratings. The negative coefficient of review length suggests that dissatisfied customers tend to express their negative emotions through longer reviews. This finding aligns with the social sharing of emotions (Rimé et al., 1992), which posits that dissatisfied customers are more likely to engage in negative word-of-mouth and elaborate their negative experiences in reviews.

On the other hand, the positive relationship between sentiment polarity and ratings supports the notion that positive sentiment expressed in reviews can influence customer satisfaction (Geetha et al., 2017; Li et al., 2012).

In Model 2, the inclusion of 20 topics identified through LDA topic modeling revealed the importance of specific aspects of the hotel experience. For instance, Topic 4, related to check-in service, displayed a negative effect on satisfaction ratings. This finding underscores the significance of efficient and smooth check-in processes in delivering a positive guest experience. It aligns with prior research highlighting the critical role of service quality and efficiency in shaping customer satisfaction (Olorunniwo et al., 2006; Parasuraman et al., 1988).

Model 3 explored the interaction between prior average rating and emotions (joy and anger). The negative interaction indicates that the influence of emotions on customer satisfaction is mitigated when the prior average rating is higher. This finding is consistent with the theory of attribution, suggesting that customers with positive prior experiences may attribute negative emotions to situational factors rather than the hotel itself (Weiner, 1985). This understanding has theoretical implications, emphasizing the need to consider customers' prior experiences when examining the impact of emotions on satisfaction.

Model 4 investigated the interaction between prior volume reviews and emotions, revealing interesting patterns. The positive interaction between prior volume reviews and anger implies that the impact of anger emotion on satisfaction increases with the number of previous reviews. This finding aligns with research highlighting the influence of social proof in shaping customer perceptions (Cialdini, 2009).

On the other hand, the negative interaction between prior volume reviews and joy suggests that as the number of previous reviews increases, the effect of positive emotions on satisfaction diminishes. This could be due to a saturation effect, where an abundance of positive reviews may lead customers to set higher expectations (Anderson, Fornell, & Lehmann, 1994). These findings highlight the interplay between emotions, social influence, and prior experiences in shaping customer satisfaction.

The presence of significant hotel dummy variables underscores the importance of considering the impact of different hotels on customer satisfaction. The significantly higher satisfaction ratings for the Baobab hotel compared to Villa del Conde and Costa Meloneras suggest that specific factors unique to Baobab contribute to its superior performance. Identifying and understanding these factors can provide valuable insights for managerial decision-making and improving customer satisfaction across all hotels. For example, Baobab may have implemented effective customer service training programs or innovative service offerings that positively impact customer experiences.

3.6 Theoretical and Managerial Implications

Based on the study's findings, the following theoretically and practically implications were performed. Most of the extracted topics were similar to the factors that have been identified in the literature as determinants of consumer (dis-)satisfaction (Dinçer & Alrawadieh, 2017; Guo et al., 2017; J. Zhang, 2019), which has some theoretical implications. Luxury hotel guests tend to be dissatisfied due to pay issues (topic 13) and a lack of amenities (topic 18) like water, towels, etc. In contrast to earlier studies, we found no issues with the extracted topics' value for money or cleanliness.

Possibly as a result of only taking into account reviews from luxury hotels. Moreover, the inclusion of emotion variables in the analysis enhances our understanding of the role of emotions in customer satisfaction. By examining the interaction effects of emotions with prior average rating and volume reviews, this study offers insights into the boundary conditions that modulate the impact of emotions on satisfaction.

For instance, our findings show that when joy embedded emotion increases, satisfaction ratings increase while anger emotions decrease them. Our results are similar to (Lee et al., 2015) in the case that when consumers rate their level of satisfaction, they are influenced by other ratings (such as the prior average rating) and the rating variance (the distribution of prior ratings). By using LDA to investigate the topics embedded in online reviews, our study adds to the body of literature about consumer satisfaction in the hospitality industry. Both the factors that affect customer (dis-)satisfaction and various customer segments were related in the extracted topics. A different approach to collect customer feedback from surveys or content analysis is topic extraction. It has practical implications by helping managers to better understand the concerns that customers talk about and make informed decisions regarding service improvement.

3.7 Limitations and recommendations for future research

The use of only three hotels within the same chain was one of this study's limitations. To check the consistency of the results, more luxury hotels should be added. Second, future research should consider other reviewer characteristics to further profile the extracted topics and explain satisfaction ratings.

In addition, recently, Ding et al. (2020), point out that the use of LDA has limitations as it cannot provide additional document-level information and it is difficult to examine the relationship between document metadata and the content of a document model (in line with Roberts et al. 2016). To overcome this limitation, the Structural Topic Modelling (STM) is an extension of LDA that allow researchers to estimate the impacts of document metadata on the prevalence of latent topics. According to Ding et al. (2020), despite STM has been applied in studies to analyse different types of reviews, only one service quality research has used STM in textual data analysis.

Specifically, this unique study refers to Korfiatis et al. (2019) research that measure service quality of airline industry by analysing online airline passengers' reviews. In the hospitality industry, the STM has only been applied to explore perceived service quality attributes in Airbnb' accommodation by Ding et al., (2020) but until now, the STM has not been applied to explore perceived service quality attributes in hotels and we suggest this recommendation for further research.

CHAPTER 4

Social influence through prior mode rating: The moderating role of geographical proximity and expertise of reviewers

4.1. Introduction

In today's era of electronic communication, the traditional reliance on personal recommendations for decision-making has been replaced by the growing trend of online reviews and ratings. Customers now have access to a wide range of perspectives and experiences from other customers due to the emergence of various online review platforms such as TripAdvisor, Yelp, and Booking. As a result, online customer reviews and ratings have become an essential source of information for the lodging industry, significantly influencing consumers' decision-making processes (Baka, 2016; Hu, Zhang, Gao, & Bose, 2019; Mariani, Borghi & Gretzel, 2019; Mellinas, Nicolau and Park, 2019). When deciding which hotel to book, potential customers often turn to online reviews for information by considering the comments written by previous customers on review pages and then they adjust their own judgements particularly if they believe those customers to be more informed or experienced (Chevalier & Mayzlin, 2006).

These reviews and ratings have become an important source of social influence (i.e., the ability of individuals to affect the behavior or attitudes of others within their social or cultural context), as they can shape the perceptions and attitudes of potential customers towards hotels (Gretzel et al., 2015; Liu & Park, 2015).

A growing body of research has shown that there is a "social influence effect" which shows that consumers' online rating behavior is influenced by their exposure to others' average review ratings (Schlosser, 2005; Li & Hitt, 2008; Moe & Schweidel, 2012; Lee, Hosanagar & Tan, 2015; Ho, Wu & Tan, 2017).

Previous research has focused on mean rating as a source of social influence (Shrihari et al., 2012; Filieri et al., 2019) but there may be other cues that people use as a social consensus indicator, such as rating mode. Mode refers to the most frequently occurring rating among a set of reviews (Köcher 2021) and it is an element very easy to discover when ratings distribution was presented on reputation platforms. For example, if a hotel has received 20 reviews, twelve reviews of which are 5-star ratings and eight are 4-star ratings, then the mode rating would be 5 stars. It can serve as a heuristic cue for what others think about a hotel and may be more salient to potential customers than mean rating (Chung 2018). As a result, consumers' online rating might be influenced by others’ ratings through the mode of rating distribution and social influence occurs at post purchase stage of consumer journey when customers assess their product or service experience (Moe et al., 2012, Schlosser 2005).

Figure 4.1 Rating distributions of mode

Hotel A

Hotel B

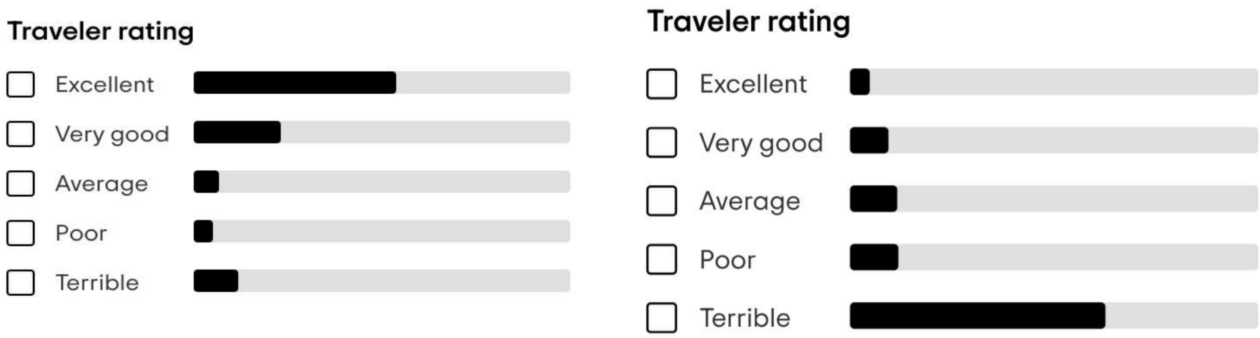


Figure 4.1 shows the distribution of mode between two hotels A & B. People may tend to use the mode as a heuristic foundation while making inferences about products, which can lead to systematic divergences in product evaluations inferred from the two distributions illustrated.

Hotel A appears to be the one that customers are more likely to select as the majority of reviews are positive, which may persuade potential customers to choose it. Hotel B, on the other hand, has a higher proportion of unsatisfied reviews, which may deter future customers from making a reservation there. Consequently, exposure to prior mode ratings can serve as an anchor for the reviewer's rating scale similar to prior average rating, and therefore be a social influence source.

Having said that, social influence is a significant source of bias in review-rating decisions, so it is beneficial to understand the variables and factors that strengthen or weaken this effect. Several studies have focused on the factors affecting reviews credibility (Banerjee et al., 2017; and Cheung et al., 2012) while other have focused on review helpfulness (Schlosser 2011; Singh et al. 2017; and Yin et al.,2014). When looking at consumer reviews more broadly, a significant amount of research is focused on how consumers react to various factors that describe how ratings are distributed, such as prior average ratings (Chevalier et al., 2006; Dellarocas et al., 2007; and Godes et al., 2004), and the volume of ratings (Liu 2006; Moe et al., 2011; and Zhu et al., 2010). Even though there has been a lot of literature devoted to understanding how consumers respond to different rating characteristics (Babic' Rosario et al., 2016), we still know very little about the consequences of additional rating features, such as the prior mode ratings.

Therefore, the goal of this study is to fill in these gaps in the literature by digging deeper into the phenomena of social influence effect at post-purchase stage on advancing our understanding of how consumers influenced by different aspects of rating characteristics such as the mode of prior ratings — or the rating score that has received the most satisfaction votes.

While previous research (Pentina et al., 2018; Zhang et al., 2014) has focused on the influence of online ratings on consumer decision making process at purchase stage (i.e., selecting from many options when they book a service or order a product after reading some reviews and checking prior average rating score), the present work investigates how individual ratings are affected by the mode at post-purchase stage.

Furthermore, the impact of the prior mode rating, as social influence, on individual ratings may depend on reviewer characteristics (Guo et al., 2016). In particular, we propose that the geographical proximity (how close or far the reviewer is from the hotel) and reviewer expertise (based on the number of reviews written by reviewers) may moderate the relationship between the mode of prior rating and subsequent rating.

According to previous research (Chen & Xie, 2008; Lu et al., 2014) the effectiveness of eWOM may be influenced by various factors, such as the source credibility, content valence, and characteristics of the product or service being reviewed. One such factor that has been found to play a moderating role in the effectiveness of eWOM is geographical distance. For instance, Schlosser et al. (2005) observed that both the substance of the review text and the geographic distance may have an impact on how eWOM affects consumer behavior.

Reviewer geographical proximity and expertise were selected as potential moderators, as they are key measures for a reviewer's influence that relate to their susceptibility to social influence, and influential reviewers are attractive targets for marketers (Hong et al., 2018; Köcher et al., 2021; Xie & Lee, 2015). By taking reviewer-side differences in response to social influence into consideration, this approach can provide valuable insights into the effect of prior mode rating on subsequent rating. We aim to fill the gap in the literature and contribute to a better understanding of the role of heuristic cues in consumer decision-making processes. Therefore, the goal of this study is to better investigate the effect of prior mode ratings on subsequent ratings and whether this effect is moderated by reviewer characteristics such as the reviewer's expertise and geographical proximity by addressing the following research questions:

1. What is effect of mode of prior ratings on post-purchase consumer satisfaction assessments (i.e., subsequent ratings)?

2. What is moderating role of the reviewer's geographical proximity on the relationship between prior mode rating and subsequent ratings?

3. What is the moderating role of the reviewer's expertise on the relationship between prior mode rating and subsequent ratings?

To investigate social influence effect and the moderators, we used a sample of 163,463 online reviews from TripAdvisor in New York City for 145 hotels. We fitted a multilevel ordinal logit model containing hotel, review text, along with other review-specific characteristics such as reviewer’s geographical proximity & reviewer expertise.

The contribution of this research is significant from both a theoretical and managerial standpoint. First, this study can prove that the mode of prior rating is a crucial factor in the social influence effect. Second, we extend to the body of literature on the decision making and the heuristic use in judgement generally by showing how people use a prior rating mode as a heuristic when making product judgments (Gilovich et al., 2002). Furthermore, this study contributes to a better insight on how the geographical proximity and reviewer expertise may change the effect of social influence on subsequent rating behavior in online reviews.

Findings of this study can be used to advise managers of reputations platforms about possible context-related factors that might increase or decrease social influence effects in reputations platforms. Results might help them to design alternative presentation of previous reviews on reputation platforms.

4.2 Literature review

4.2.1 Social influence and rating distributions

Online reviews have become increasingly important to consumers in the lodging industry as they rely on them to make informed decisions (Hu et al., 2019). Prior research has shown that customer decision-making is significantly influenced by online reviews and ratings, as customers are exposed to various heuristic cues, such as the average rating, dispersion in ratings, and the quantity of prior reviews (Filieri & McLeay, 2014; Gretzel et al., 2015; Li et al., 2020; Li et al., 2019; Gavilan et al., 2018). Furthermore, conformity issues (i.e., when people conform to the thoughts and beliefs of others, even if those beliefs conflict with their own personal preferences) can also play a role in influencing people's decision-making processes (Cai et al., 2009).

Studies have found that the posting of early positive ratings can lead to the posting of negative ratings, creating a differentiation and correction effect (Liu & Park, 2015; Moe & Trusov, 2011; Dellarocas, 2003). However, people's judgment processes may be influenced by heuristics, which can lead to an incorrect weighting of informational cues, such as the anchoring heuristic (Tversky & Kahneman, 1974). This heuristic leads people to depend extensively on the initially acquired information, even if it is unrelated to the task at hand (Köcher et al., 2021). As a result, earlier ratings or comments can lead to herd behavior, in which people tend to arrive at the same conclusion (Cicognani et al., 2016). Given the importance of online reviews, various measures can be used to illustrate the position of customer ratings, such as frequency of rating scores, locational measures (like mean and mode), variability-based measures (like standard deviation and variance), and shape-based measures (like skewness and kurtosis) (Köcher et al., 2021).

Previous studies have focused on the impact of review valence and review volume on sales rates investigating social influence effects at purchase stage of consumer journey (Chevalier et al., 2006; Chintagunta et al., 2010; Clemons et al., 2006; Dellarocas et al., 2007; Godes et al., 2004; Liu, 2006; Moe & Trusov, 2011; Sun, 2012; Zhu et al., 2010). Thus, according to previous findings, at the purchase stage, companies need that their products or service present positive online reviews to positively influence purchase decisions of their prospect customers. At post-purchase stage, companies need positive feedbacks and ratings in reviews from their customers, based on real customer experience (i.e., not biased), and therefore not influenced by others (i.e., absence of social influence). Therefore, companies want social influence at the purchase stage, but they do not want social influence effect at post-purchase stage for products or service assessment.

Post-purchase service and product assessments are based on customer experience which is very subjective and rely on expectations and experience with contextual factors such as customers-service employee interactions, customer characteristics, product/service characteristics, weather when writing a review (Xue et al., 2020) and even reputation platform settings such as graphical distribution of previous ratings, indicating badge for reviewer's expertise or previous reviewers' location. Similarly, the review mode can affect online ratings and purchase behavior. A recent study by Köcher et al. (2021) investigated the role of the mode in how consumers interpret various aspects of online rating distributions.

Our study differs from Köcher et al. (2021) by investigating how mode of prior ratings influence consumer satisfaction ratings at post-purchase stage of the consumer journey. Moreover, this study investigates boundary conditions of social influence effect at post-purchase stage through mode of prior rating. Based on previous findings, people may tend to rate higher when they see the location of the mode ratings above the prior average ratings, or they may give the opposite direction of others' average rating, implying a positive or negative effect of social influence on rating behavior (Adomavicius et al., 2013). Individuals often estimate unknown values by starting with a prominent anchor value and adjusting from that anchor until a satisfying value is obtained (Tversky & Kahneman, 1974). Similarly, a reviewer may start with the location of the mode rating anchor and then apply an anchoring-and-adjustment heuristic to a rating choice based on their (dis-)confirmation of their expectations.

In conclusion, despite extensive research on consumers' reactions to prior average rating at purchase stage, information regarding the prior mode rating as well as its moderators at post-purchase is still lacking. Therefore, the purpose of this research is to show that the mode of prior rating exists as a social influence effect in online reviews. The mode may be frequently used by people as a heuristic when making rating decisions at post-purchase stage because it is very easy to see graphically when consumers expose to distribution of prior ratings. More specifically, since the mode receives the highest frequency of positive ratings, it might be more effective than any other factor at drawing people's attention therefore, it possibly considers the most important cues when forming an overall opinion (Ibrekk et al., 1987; and Weber et al., 1997).

Therefore, we state that the mode rating might has a direct effect on subsequent ratings.

Hypothesis 1: The mode of prior ratings has a positive effect on subsequent rating.

4.2.2 The moderating role of reviewer's geographical proximity

Marketers are debating whether geographical location (i.e., is the location of the reviewer in respect to the location of the good or service being evaluated) plays a part in eWOM effectiveness as the availability of location data increases. Some marketers believe that eWOM coming from a longer distance may be more broadly accepted, increasing its usefulness and plays a part in eWOM effectiveness as the availability of location data increases. Some marketers believe that eWOM coming from a longer distance may be more broadly accepted, increasing its usefulness and informativeness to customers (eMarketer, 2019).

However, other marketers argue that eWOM coming from a more localized population is more specific and relevant, making it more reliable (Backaler, 2018; Sharma, 2018), whereas, others claim that consumers' perceptions of the value of eWOM are not impacted by distance (Quinn, 2018). Geographical distance may still have an impact on online consumer behavior and WOM due to various factors, as pointed out by Goldfarb and Tucker (2019). These factors include the existence of location-specific goods and economic costs such as shipping, contracting, monitoring, enforcement, travel, and inconvenience, as indicated in studies by Adamopoulos, Ghose, and Tuzhilin (2021) and Hortaçsu, Martínez-Jerez, and Douglas (2009). Furthermore, local user preferences and spatially correlated social ties, as well as limited regional availability of alternative product options, can also result in similarities in product adoption among nearby consumers, as demonstrated by studies by Bell and Song (2007), Ma, Krishnan, and Montgomery (2014), and Meyners et al. (2017). These reasons suggest that geography may still limit the effectiveness of eWOM. The emergence of new communication technologies that are not influenced by geographic distance has not altered this relationship (Mok, Wellman, and Carrasco, 2010). For example, Twitter links, social networking ties, and email traffic all decrease with distance (Lee, Scherngell, and Barber, 2011; Takhteyev, Gruzd, and Wellman, 2012). While it is well-established that the frequency of interaction and communication depends on geographic proximity, it remains unclear whether social influence is more likely to occur and stronger when individuals are geographically close to products or services.

This study aims to explore this issue by investigating whether geographical distance proximity is significantly associated with the effect of rating behavior. Specifically, we investigate whether the geographical proximity of reviewer makes a difference in social influence effect at post-purchase stage.

We draw our argument based on social identity theory by (Turner and Tajfel, 1986). This theory suggests that a geographic area can be considered part of an individual's social identity, leading them to perceive geographically close individuals as more similar to themselves (Huddy and Khatib, 2007; Turner and Tajfel, 1986). According to social identity theory, an individual's social identity is formed by their sense of belongingness to a community (Ashmore, Deaux, and McLaughlin-Volpe, 2004; Tajfel, 1974). The shared social identity between the message sender and recipient can act as a heuristic to guide judgments and actions (Chaiken, 1987; Chaiken and Maheswaran, 1994), giving social identity a strong influence on human behavior (e.g., Atefi et al., 2018; Jenkins, 2014; Shang, Reed, and Croson, 2008). Proximity might be an important factor when deciding which hotels to choose while evaluating them, potential guests may be more likely to believe and be swayed by reviews of others who are close to them in terms of location. That is, if a reviewer locates in the same country where the hotel is situated, for example US, reviewer might be more open to social influence because they would expect more customers from the same country (e.g., US) stayed in the same hotel, therefore he will be more likely to be affected by previous rating. So, we postulate that the geographical proximity will moderate the association between the mode of the initial evaluation and later ratings, making the social influence effect larger for nearby customers.

Therefore, we hypothesize that:

Hypothesis 2: The relationship between mode of prior rating and subsequent rating is moderated by reviewer's geographical proximity where the relation is stronger for reviewers located close to the hotel.

4.2.3 The moderating role of reviewer's expertise

With the rise of various review platforms such as TripAdvisor, Yelp, and Booking, customers can easily share their opinions and experiences with others. However, electronic word of mouth (eWOM) is often communicated by anonymous users, which reduces its credibility compared to traditional word of mouth (Park et al., 2007). Therefore, identifying the source credibility of eWOM is particularly challenging for readers. To address the issue of anonymity, review websites offer ranking systems that allow readers to establish social networks of reviewers and rank them based on the quantity and quality of their reviews. This ranking system increases the source credibility of the reviewer, as it demonstrates their level of expertise and socialization within the online community (Köcher et al., 2021; Park & Lee, 2009). Nonetheless, studies indicate that the information provider has a significant influence on consumer preference and choice (Herr, Kardes, & Kim, 1991).

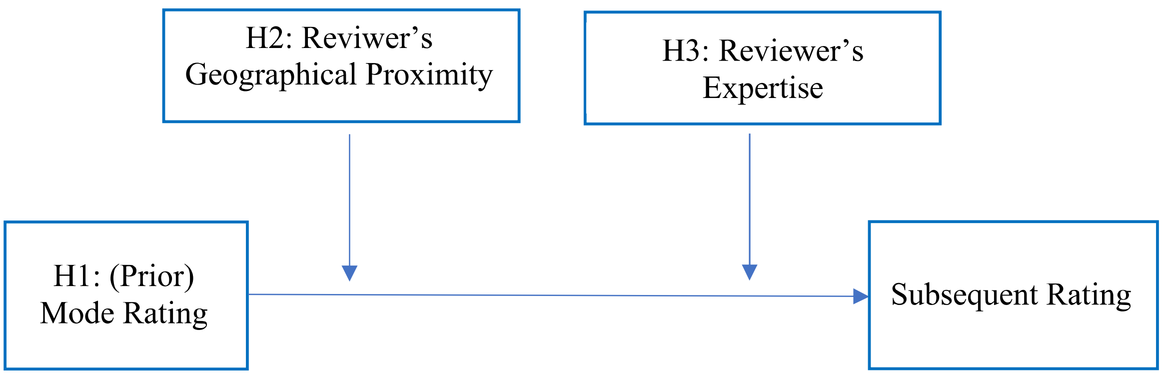
Recent studies have suggested that the relationship between review order and review helpfulness may be moderated by reviewer expertise (Zhu et al., 2020; Hong et al., 2018). Reviewer expertise can be defined as the number of reviews the reviewer has written for a particular product, and reviewers with a high level of expertise are less dependent on others for product-related information and are less susceptible to normative influence from prior reviewers (Hu et al., 2021; Xie & Lee, 2015). This suggests that the subsequent rating from a more expert reviewer is less likely to be affected by its prior mode rating. Conversely, less expert reviewers are more likely to seek information and approval from prior reviewers and comply with their opinions as an ego-defence mechanism to avoid social disapproval (Hu et al., 2021; Xie & Lee, 2015).

Therefore, we hypothesize that the relationship between prior mode rating and subsequent rating is moderated by reviewer expertise, such that the relationship between … should be weaker for reviewers who have a higher level of expertise, consequentlythe effect of prior mode rating on subsequent rating is stronger for less expert reviewers than for more expert reviewers. Thus,

Hypothesis 3: The relationship between prior mode rating and subsequent rating is moderated by reviewer expertise where the relation is weaker for reviewers who have higher level of contributions.

The research framework was shown in fig. 4.2

Figure 4.2 The Conceptual Framework



4.3 Methodology

4.3.1 Data

The goal of this study is to determine whether the effects of the mode could be measure in actual real data. Therefore, we collected our data from TripAdvisor platform, one of the most popular online sources for hotel comments (comScore, Inc., 2016). We chose hotels in New York City, a leading tourism city that accommodates many domestic and international tourists annually, to include hotels with diverse price scales and to guarantee an adequate number of reviews for each hotel included. Moreover, TripAdvisor is among the hotel review websites that academics most frequently study.

The study sample included 163,463 online reviews of 145 New York City hotels on TripAdvisor between January 2015 to July 2022.

To ensure an adequate number of reviews for each hotel included, we selected hotels in New York City, a popular tourist destination that hosts numerous domestic and foreign tourists each year. Consumers' online hotel ratings, the review text, and the review helpfulness were all included in the data.

The collected data was analysed using Python software and each review is analysed to calculate its polarity. This method takes into account the strength of the sentiment, with values ranging from -1 (extremely negative) to 1 (extremely positive). VADER (Valence Aware Dictionary and sEntiment Reasoner) a lexicon and straightforward rule-based model for sentiment analysis, is used to achieve this (Hutto & Gilbert, 2014).

4.3.2 Variables

The dependent variable was the reviewer's rating (Y_{ijt}) on a scale of one (terrible) to five (excellent) (Li et al., 2019, 2020). The analyses included a number of review-text variables, and reviewer variables.

Independent variables

Social influence was operationalized as the mode of previous ratings of the hotel which means the highest frequency of satisfaction rating. Thus, the independent variable "social influence" was measured as the prior mode rating ($Mode_{ij(t-1)}$) and it could have values between 1 and 5. Note that considering the review rating distribution is not a symmetrical distribution, mode of prior ratings will be very likely different than prior average rating despite they are both locations parameters of a distribution. Therefore, it worths to study what happens to social influence when mode and average prior rating differs a lot.

Moderators

Geographical proximity variable was used as a moderator to understand how it influences the effect of mode of prior ratings on subsequent rating.

This variable was extracted from TripAdvisor as the location of the reviewer. Since all selected hotels in New York, geographical proximity was measured as whether a reviewer in USA or not. We generated a dummy variable where the value is 1 when the reviewer location is in the United states and 0 when the reviewer location is out of the U.S.

Reviewer expertise variable was used as a second moderator to examine the effect of mode of prior ratings on subsequent rating. This variable was extracted from TripAdvisor as the reviewer contribution where higher number of reviews represent higher contribution.

Control variables

Since long reviews imply greater effort by the reviewer compared to short reviews and are more unfavourable than short reviews (Forman, Ghose & Wiesenfeld, 2008), we controlled for review length, which was assessed by the logarithm of the number of words in the review (Poncheri et al., 2008; Li et al., 2020). The natural logarithm transformation was used to normalize this measurement because the range of number of words was highly scattered.

The *polarity* score is ranges from -1 (most extreme negative) to 1 (most extreme positive), higher score implies that the text has a more positive sentiment. A score of 0 indicates that the sentiment is neutral (Geetha et al., 2017).

Table 4.1 describe the variables in our model and Table 4.2 illustrates the correlations between the variables. Figure 4.3 shows the distribution of hotel ratings. Similar to previous findings, the distribution of hotel ratings is very skewed, with most of them being 5 or 4. (Mariani & Borghi, 2018). The majority are either "5: very satisfied" (54%), or "4: satisfied" (26%), which is consistent with prior literature. A small number are "3: so-so" (10%), "2: dissatisfied" (5%), and "1: very dissatisfied" (4%).

Figure 4.3 The distribution of ratings in reviews

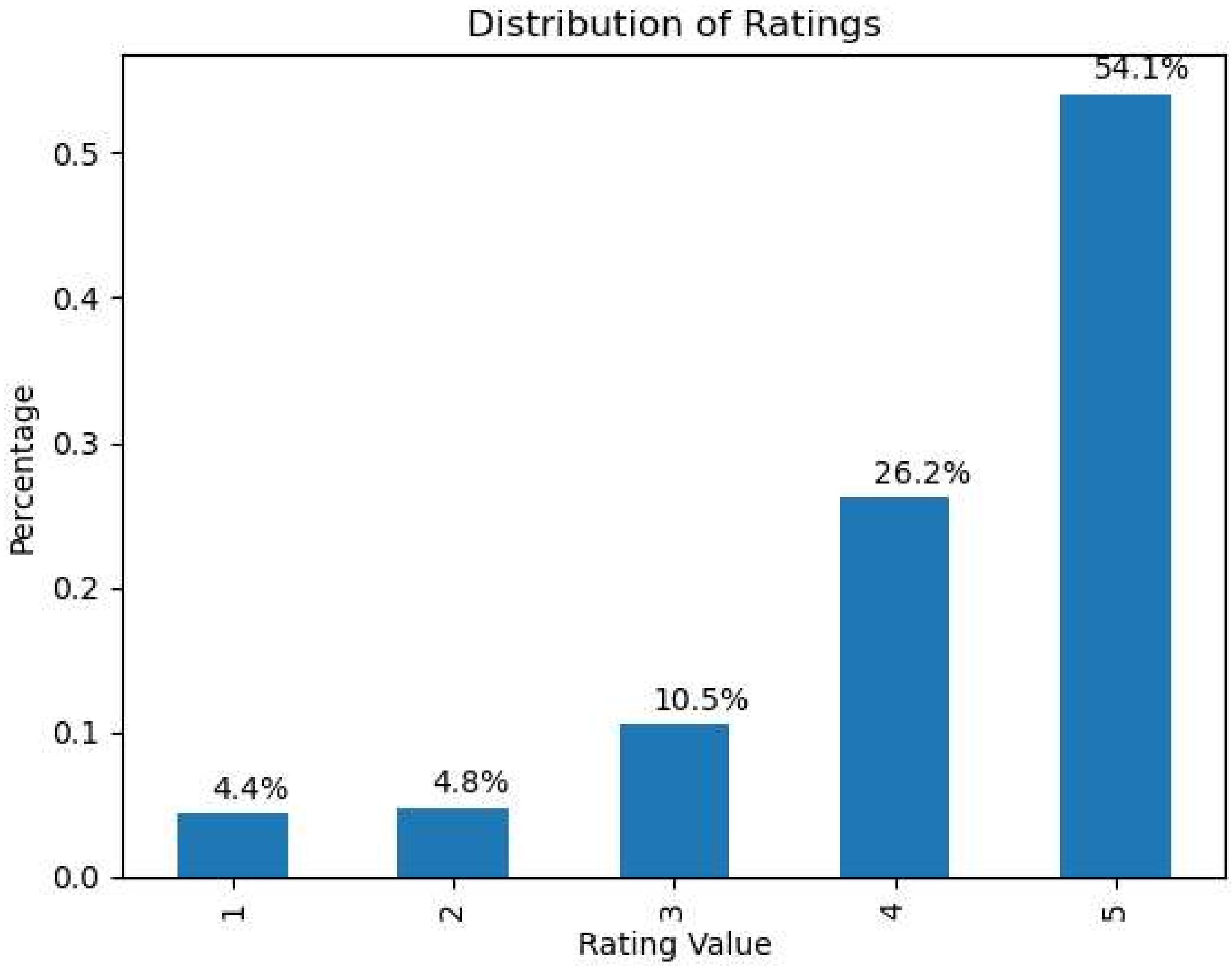


Table 4.1 Variable description

Variables	**Definition**
DV: Rating (Y_{ijt})	Review rating in review *i* for hotel *j* at time *t*
IV: $Mode_{ij(t-1)}$	The mode of prior ratings for hotel *j* at time *t-1*
Moderators: Geographical $Proximity_{ijt}$ Reviewer $Expertise_{ijt}$	 Dummy variables where 1 is in the US and 0 is out of the US. Reviewer contributions based on the number of reviews

Control variables:	
$Length_{ijt}$	The number of words in review i.
$Polarity_{ijt}$	Sentiment score between -1 and 1 in a review text i.

As we can see from Table 4.2, there is no high correlation between independent variables.

Table 4.2 Spearman correlation between variables (N = 163,463)

Variables	Mean (SD)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Rating	4.2 (1.06)	1.000						
(2) Mode	4.8 (0.4)	0.189*	1.000					
		(0.000)						
(3) PriorAvgRating	4.2 (0.3)	0.326*	0.558*	1.000				
		(0.000)	(0.000)					
(4) GeographicalProximity	0.5 (0.5)	0.047*	0.052*	0.051*	1.000			
		(0.000)	(0.000)	(0.000)				
(5) ReviewerExpertise	107.8 (1025.1)	-0.007*	-0.001	-0.002	0.008*	1.000		
		(0.007)	(0.714)	(0.343)	(0.002)			
(6) Polarity	0.8 (0.4)	0.600*	0.096*	0.198*	0.005*	0.003	1.000	
		(0.000)	(0.000)	(0.000)	(0.037)	(0.193)		
(7) Length	497.4 (221.1)	-0.190*	-0.048*	-0.045*	-0.045*	0.039*	-0.063*	1.000
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	

*** *p<0.01*, ** *p<0.05*, * *p<0.1*

4.3.3 Measures

In this study, we explored the geographical proximity and reviewer expertise as a moderating factor that influence the relationship between prior mode rating and subsequent ratings. We calculated the number and average rating for each hotel.

For control variables, we also calculated the review sentiment polarity, review length, and considered them into our analyses. Since there are multiple reviews per hotel and the dependent variable is ordinal in nature, a multilevel ordinal logit model was preferred.

We z-standardized the variables that we used in our model before the analysis to make it possible to compare the magnitude of effects easily. The unit of analysis is the review. A review rating y_{ijt} can take values among 1 and 5, which is the rating given by reviewer i for hotel j at time t. There is total 163,463 reviews and 145 hotels. Y^*_{ijt} indicates the latent variable of reviewers' hotel evaluation. For instance, y^*_{ijt1} is the probability that hotel j is rated 1, and y^*_{ijt4} is the probability that hotel j is rated less than or equal to 4.

The following multilevel ordinal regression model was fitted using standardized variables:

Yijt = β0 + β1(zPriormodeijt) + β2(geo_proximityij) + β3(rev_expertiseij) + β4(zPriormodeijt * geo_proximityij) + β5(zPriormodeijt * rev_expertiseij) + β6(zPolarityijt) + β7(zlengthijt) + εijt

where:

- Y_{ijt} is the ordinal response variable for the reviewer on the review of the hotel.
- β_0 is the intercept term.
- $zPriormode_{ijt}$ is the standardized mode of prior rating for the reviewer.
- $geo\_proximity_{ijt}$ is the measure of geographical proximity for the reviewer.
- $rev\_expertise_{ij}$ is the measure of reviewer expertise.
- $zPolarity_{ijt}$ is the standardized measure of sentiment polarity for the reviewer.
- $zlength_{ijt}$ is the standardized measure of review length for the reviewer.
- εijt is the error term.

The model includes two interaction terms: zPriormodeijt * rev_expertiseij and zPriormodeijt * geo_proximityij. These terms allow for the possibility that the effect of zPriormodeijt on Yijt may depend on the values of rev_expertiseij and geo_proximityij.

4.3.4 Results

Table 4.3 presents the outcomes of the proposed research model. Model 1 is the model with only control variables. Model 2 is the base model, which includes the main effects of the mode of prior rating and control variables. Model 3 examined the interaction effect of geographical proximity with the mode of prior rating on subsequent rating. Model 4 investigated the interaction effect of reviewer's expertise with the mode of prior rating and subsequent rating.

Models 2 is used to test our first hypothesis while models 3&4 are used to test the second and third hypotheses. Therefore, Models 2, 3 and 4 are thus utilized to describe the final estimation results below. All control variables were significant in Model 1. We first interpreted the main effects of the mode of prior rating and moderating variables because they constituted the basis for our hypotheses. According to Model 2, mode of prior rating has a significant positive effect on subsequent ratings ($\beta = 0.117$, $p < 0.001$) which support our primary hypothesis (H1).

Thus, when prior mode rating increases one unit, satisfaction rating increases 12%. Polarity has a positive effect on subsequent ratings ($\beta = 1.134$*)*; that is, increased positive sentiment increased ratings. Review length has a negative effect ($\beta = -0.364$) on subsequent ratings indicating that dissatisfied hotel guests wrote longer reviews.

Based on Model 3, The interaction of geographical proximity has a positive effect (β = 0.0491) on the relationship between prior mode rating and subsequent rating, this means that the social influence effect through prior mode rating is stronger for reviewers located close to the hotel (i.e., American reviewers compare to non-American ones) which also supports our second hypothesis (H2). Model 4 shows the moderating effect of reviewer expertise which has a negative effect (β = - 0.0000194) on the relationship between prior mode rating and subsequent rating, this means that the social influence effect via prior mode rating is weaker for reviewers who have higher level of contributions (number of written reviews). Therefore, H3 is supported by our results. A summary of the hypotheses results is presented in Table 4.4. Thus, the positive relation between the mode of prior ratings and subsequent ratings is weaker for reviewers who have a higher level of contributions.

Overall, the results suggest that the mode of prior rating as a social influence is an important predictor of subsequent ratings, and that the relationship may be influenced by the reviewer's geographical proximity and expertise.

Table 4.3 TripAdvisor (standardized values used)

	M1	M2	M3	M4
Length	-0.364***	-0.364***	-0.360***	-0.363***
Polarity	1.133***	1.134***	1.135***	1.134***
Mode of PriorRating		0.117***	0.0953***	0.119***
GeographicalProximity			0.216***	
GeographicalProximity*Mode of PriorRating			0.0491***	
ReviewerExpertise				-0.0000165**
ReviewerExpertise* Mode of PriorRating				-0.0000194**
cut1	-4.520***	-4.550***	-4.445***	-4.552***
cut2	-3.268***	-3.298***	-3.194***	-3.300***
cut3	-1.742***	-1.772***	-1.667***	-1.774***
cut4	0.0778	0.0513	0.160**	0.0495
var(cons)	0.492***	0.381***	0.373***	0.381***
N	162882	162882	162882	162882

Table 4.4 Summary of Hypotheses-Testing Results

Hypothesis	Empirical support
Hypothesis 1: The mode of prior ratings has a positive effect on subsequent ratings.	√
Hypothesis 2: The relationship between mode of prior ratings and subsequent ratings is moderated by reviewer's geographical proximity where the relation is stronger for reviewers located close to the hotel.	√
Hypothesis 3: The relationship between mode of prior ratings and subsequent ratings is moderated by reviewer expertise where the relation is weaker for reviewers who have higher level of contributions.	√

4.4 Discussion

The objective of this study is to enhance the understanding on what conditions may change the impact that social influence, by way of mode of prior ratings, has on subsequent hotel review ratings by learning how consumers interpret various aspects of online rating distributions based on the mode of previous ratings. This paper employed text analytics to scrape reviews from TripAdvisor website.

Moreover, we included a variety of meaningful control variables, such sentiment polarity and review length. Our study offers convincing empirical support for social influence effect, in terms of mode of prior ratings. Our findings show that—even when controlling for the effects of review length and sentiment polarity—subsequent ratings were positively affected by the mode of prior ratings. When people read the comments written by previous customers on review pages, they might adjust their own evaluations in accordance (i.e., when they see the majority of previous reviews have high ratings, they adjust their rating and increase their satisfaction ratings) As a result, social effects may have an impact on how consumers rate products in online reviews, not only with prior average rating (Moe et al., 2012, Schlosser 2005) but also with prior mode ratings. As a social consensus indicator, the mode of prior ratings can be used as a heuristic cue which can be used to help people evaluate reviews before making their final decisions. In addition, a study by Hong et al. (2018) introduces three types of potential moderators to examine the relationship between review order and review helpfulness: reviewer characteristics, review characteristics, and situational factors. These moderators are selected because they can help to better understand the underlying motives of reviewers and their response to social influence (Shasha et al., 2017). The study highlights that failing to account for these differences may lead to biased estimation of the social influence effect of prior reviews on focal reviews (Kuan et al., 2014).

Our results demonstrated that the positive effect of the mode of prior ratings and subsequent ratings is stronger when the reviewer geographical proximity is closer to the hotel, which is consistence with our hypothesis (H2).

When a reviewer is located in the same country of the hotel, such as the USA, they may be more susceptible to social influence because there will be more prior USA guests who have stayed at the same hotel, and their evaluations and ratings will be seen as more comparable. Therefore, the proximity to the hotel will attenuate the association between the mode of the initial evaluation and later ratings, making the social influence effect larger for nearby customers. In addition, our results show that the positive effect of the mode of prior ratings and subsequent ratings is weaker for reviewers who have have higher level of expertise (in terms of the written reviews), which is also consistence with our hypothesis (H3). This suggests that the subsequent rating from a more expert reviewer is less likely to be affected by its prior mode rating. Conversely, less expert reviewers are more likely to seek information and approval from prior reviewers and comply with their opinions as an ego-defense mechanism to avoid social disapproval (Hu et al., 2021; Xie & Lee, 2015).

4.5 Theoretical implications

This research provides a variety of contributions. It is one of the few studies in the literature on hospitality and tourism that demonstrates how the mode of prior ratings, as a social influence, affect the rating behaviors in hotel reviews. Consumer may be socially influenced by many factors such as others’ ratings. More importantly, the effect that social influence has on subsequent ratings will change depending on the reviewer expertise and geographical proximity. While previous research on social influence has mainly concentrated on the effects of average ratings, review helpfulness, and review sentiments (Jain et al., 2021; Gao et al., 2017; Hu et al., 2014), our study is the first to identify the prior mode of rating as a social influence factor and suggest that the prior mode of rating can have a significant impact on subsequent ratings.

Secondly, our study shows that the effect of mode of prior rating is moderated by reviewer characteristics such as the geographical proximity and reviewer expertise. This shows the importance of taking the reviewer characteristics into account when studying social influence in the online context. Therefore, the current study contributes to the literature of social influence studies (Ma et al., 2014; Schlosser, 2005; Li & Hitt 2008; Moe & Schweidel, 2012; Lee, Hosanagar & Tan, 2015; Ho, Wu & Tan, 2017) in terms of mode of prior ratings by providing empirical analysis on how the social influence works and how it can be moderated using reviewer characteristics.

Finally, this study contributes to the literature on the heuristic judgement and decision making by demonstrating the use of different cues in the evaluation process (Gilovich et al. 2002). Individuals frequently estimate unknown values by starting with a prominent anchor value and adjusting from that anchor until a satisfying value is obtained (Tversky & Kahneman, 1974). Then, an expert reviewer (ie. reviewers who wrote many reviews) influenced less by other ratings. Our results show that the mode of prior rating might be a heuristic framework when people evaluate a product, and this effect depends on some characteristics such as the more with proximity to the hotel and less with reviewer expertise.

4.6 Managerial contributions

Although the main objective of reputation platforms is to influence potential customers to buy a product or a service based on customer reviews, social influence, particularly through mode of prior ratings, is not always advantageous for a company as the mode might cause some customers to alter their ratings and not give an honest rating after being exposed to prior ratings by others.

In order to investigate the role of mode of prior ratings, this study has investigated the moderators that can strength or weaken the social influence. This study will help the industry evaluate products more accurately. The managerial implications of our findings suggest that reputation platforms should be aware of the potential impact of social influence on subsequent ratings and consider alternative ways of presenting previous reviews to consumers. For example, they could highlight reviews from reviewers who have a history of providing balanced and objective reviews, rather than just presenting the mode of prior ratings.

The results of this study can be utilized to advise reputation platforms' managers about potential context-related aspects that could strengthen or weaken the impacts of social influence on those platforms. Results might help in the development of alternate reputation platform presentations for previous reviews.

Additionally, proximity to the hotel and reviewer expertise should also be taken into consideration when designing reputation platforms. Reputation platforms could prioritize reviews from nearby customers or more expert reviewers to provide a more accurate representation of the hotel's quality. In other words, marketers are better able to predict how the geographical proximity and reviewer expertise will affect customers' decsion making and, ultimately, how well their products will be sold. Based on our finding, managers should convince customers to write more reviews so they can be more expert and less influenced by other ratings. Moreover, hotel managers should observe hotel ratings and mode over time for USA and Non-USA and if they see decline for Non-USA ones, they should do actions (increase in service quality) so giving more attention to Non-USA customers if the hotel is in USA.

4.7 Limitations and future research

Despite the fact that our research offers important perspectives on how customers react to the mode of prior ratings, it has some limitations that suggest exciting new directions for future study. Firstly, we did not take into consideration the issue of fake comments or selection bias in reviews. According to O'Connor, 2010 there may be fake reviews on TripAdvisor.

Moreover, future studies can take into account the cross-cultural study on the moderator of social influence effect. Second, this study is only considering data from TripAdvisor which may not generalize the results. Therefore, future research should consider different platforms to measure the mode of prior ratings. Moreover, experimental research could be interesting to explain social influence mechanism with different platform presentations such as graphical representing prior average rating and mode (so no numbers to be used by as an anchor) and observe what happens to social influence effect.

CHAPTER 5

GENERAL CONCLUSIONS

5.1 Research Conclusion

The purpose of this dissertation is to use text mining and sentiment analysis techniques to better understand the social factors that affect consumers' online review behavior. This dissertation has investigated the role of prior average ratings and the prior mode on subsequent ratings by proposing that the emotions previously felt by consumers during their stay and expressed in a review text and the deviation mode can be a possible moderator of the social influence effect on consumer rating behaviours. Moreover, this dissertation has examined the product/service failures and brand positioning problems, for instance, chain hotels using topic modeling analysis.

Study 1 utilises a text mining technique based on TripAdvisor's online review data to examine how social influence (i.e., the prior average rating) affects subsequent ratings and to what extent the reviewers' expressed emotions (i.e., joy and anger) and hotel characteristics (i.e., price category and affiliation) can affect the relationship between social influence and subsequent ratings. To investigate social influence effect and moderators, we used a sample of 65215 online reviews from TripAdvisor and validated the findings with 67534 reviews from Expedia for 169 hotels in New York City.

The following research questions were examined in this study: (1) What impact do emotions in review texts have on the ratings of subsequent hotels? (2) How do product characteristics affect the individual rating behaviour?

Study 2 analyses the influence of topics and other review-specific variables, such as emotions embedded in reviews, on satisfaction ratings for luxury chain hotel reviews. This is done by extracting topics from review texts using topic modeling and sentiment analysis. We used a sample of 8,376 online reviews from TripAdvisor between 2002 to 2019. This study examined the following research questions: (1) What topics emerge in online reviews from guests of luxury hotels? (2) How do these issues, along with other review elements like the sentiments expressed in the review text, impact satisfaction scores? (3) How do these topics vary across three luxury hotels?

Furthermore, Study 3 explores how consumers interpret different aspects of online rating by the mode of prior ratings — or the highest score that has received the most satisfaction votes. To investigate social influence effect and moderators, we used a sample of 163,463 online reviews from TripAdvisor in New York City for 145 hotels. This study investigated the following research questions:

(1) What is effect of mode of prior ratings on post-purchase consumer satisfaction assessments (i.e., subsequent ratings)? (2) What is moderating role of the reviewer's geographical proximity on the relationship between prior mode rating and subsequent ratings?

(3) What is the moderating role of the reviewer's expertise on the relationship between prior mode rating and subsequent ratings?

The dissertation's results can be summarized as follows:

- ***First,*** subsequent ratings increased when prior average ratings increased as well. As a result, the social influence effect was proved to be favorable in the context of hotels in terms of prior average ratings.
- ***Second,*** when anger was the emotion expressed in reviews, the influence of previous average review ratings on subsequent ratings was stronger; however, this influence was weaker when a customer wrote a review that included joy. When customers see high prior average ratings after a positive consumer experience, they consequently lower their ratings, which could result in a downward trend for a hotel over time. However, reviews with anger emotions might adjust and increase their subsequent ratings due to a reduction in their negative feelings and to feel better.
- ***Third,*** while the impact of prior average ratings on subsequent ratings was stronger for budget hotels, the effect that social influence had on subsequent ratings did not change for luxury hotel guests and independent hotels.
- ***Fourth,*** when a review mentioned earlier reviews or reviewers, the positive influence that those reviews had on subsequent ratings was weaker (less positive).
- ***Fifth,*** subsequent ratings decreased as review length increased, this means that we people have a negative experience they tend to express their feeling more by writing longer reviews. Additionally, polarity had a positive impact on subsequent ratings.
- ***Sixth,*** when controlling for the effects of review length, and polarity—subsequent ratings were positively affected by the prior mode ratings. This means that the mode of prior rating, as a social influence factor, has a direct effect on subsequent ratings.

- ***Seventh,*** the results demonstrated that the positive effect of the prior mode ratings and subsequent ratings is moderated by reviewer's geographical proximity where the relation is stronger for reviewers located close to the hotel.
- ***Eighth,*** the results showed that the positive effect of the prior mode ratings and subsequent ratings is moderated by reviewer expertise where the relation is weaker for reviewers who have higher level of contributions.
- ***Ninth,*** customers who previously had a good experience with the hotel may be less affected by sentiments and emotions during their current stay. In contrast, customers who have had a bad experience in the past might be more likely to experience an emotional impact during their current visit.
- ***Tenth,*** the impact of anger emotion on customer satisfaction increases with the number of previous reviews (customers are more likely to be influenced by negative feelings stated in online evaluations when there are plenty of prior reviews). Contrarily, when there are more previous reviews, the impact of the joy emotion on customer satisfaction is smaller (i.e., customers are less likely to be influenced by favorable emotions stated in other customers' reviews when there are numerous prior reviews).

5.2 Research Contributions and Implications

By providing new theoretical perspectives, this dissertation adds to the growing body of literatures on both hospitality and marketing in general. The empirical findings provide important managerial implications for the hospitality industry and online review societies. First, we tested the social influence effect on subsequent ratings in online consumer reviews. Prior reviews written by other customers seem to have a social influence on online reviews.

A growing body of literature disputes that review rating environments, such as prior average review ratings and variance in prior ratings, have an impact on consumers' online review behaviour (Ho, Tan, & Wu, 2017; Lee, Hosanagar, & Tan, 2015; Li & Hitt, 2008; Moe & Schweidel, 2012). This suggest that online reviews behaviors might be influenced by social factors. Using online big data from TripAdvisor, this dissertation proposed a theoretical framework on how consumer online review is socially influenced. More importantly, the effect that social influence has on subsequent ratings will change depending on the emotions expressed and the hotel's characteristics. Therefore, our findings contribute to the literature on social influence and online reviews.

From the managerial standpoints, the findings of this dissertation help practitioners to better understand how subsequent ratings are socially influenced by prior mean and mode ratings and the important topics that consumers are talking about on online reviews.
As a result, managers can better understand the moderators that can reduce or increase social influence which would benefit the industry in terms of accurate product evaluations and improving the service quality.

Second, this dissertation employs a text mining strategy to evaluate the moderating effects of emotions embedded in review texts and product characteristics. Emotions embedded in reviews have been used mostly to explain review helpfulness (Yin et al., 2014; Felbermayr & Nanopoulos, 2016; Ren & Hong, 2019; Chen & Farn, 2020), but not individual subsequent ratings.

Therefore, the current dissertation adds to social influence literatures in terms of prior average ratings by examining hotel characteristics and two of the most meaningful emotions in reviews (Ma et al., 2014; Schlosser, 2005; Li & Hitt 2008; Moe & Schweidel, 2012; Lee, Hosanagar & Tan, 2015; Ho, Wu & Tan, 2017).

Third, this dissertation uses topic modelling and sentiment analysis of big data to determine the factors that affect customer satisfaction. By scrapping online reviews written by visitors to three luxury hotels in Spain, we used an exploratory approach to investigate the factors that affect the customer's satisfaction. According to the study's findings, the majority of the topics that were extracted shared characteristics with the elements that have been identified in the literature as determinants of consumer (dis-)satisfaction (Dinçer & Alrawadieh, 2017; Guo et al., 2017; J. Zhang, 2019). Luxury hotel guests tend to be dissatisfied due to pay issues and a lack of amenities like water, towels, etc. In contrast to earlier studies, we found no cleanness or cost effectiveness issues in the extracted topics.

Forth, the use of various cues in the evaluation process is another way that this dissertation adds to the body of literature on heuristic judgement and decision making (Gilovich et al. 2002). People frequently begin their estimation of unknown values with a prominent anchor value and adjust from that anchor until they arrive at a satisfying value (Tversky & Kahneman, 1974). Then, a reviewer can start with the prior mode rating anchor and then applies an anchoring-and-adjustment heuristic to a rating choice based on their (dis-)confirmation of their expectations. Our results show that the prior mode of rating can be served as a heuristic framework when people evaluate a product. Therefore, the prior mode aligns with other previously heuristics (Tversky et al., 1973 and Tversky et al., 1974).

Additionally, this dissertation provides marketers and managers with valuable managerial implications for the manipulation of online reviews and its effects. This dissertation shows that customers who have a negative experience are more likely to write longer reviews with stronger polarity, which could be more harmful to a business's reputation and long-term profitability. Furthermore, this dissertation offers significant managerial implications for those who manage online review systems.

To investigate the role of embedded emotions as a proxy for guest experience indicators and product characteristics in online consumer hotel reviews, this dissertation has investigated the moderators that can reduce or increase social influence. Therefore, the managers of reputation platforms can use our findings to adopt new methods of summarizing ratings from previous reviews or to avoid exposing reviewers to average ratings while they are writing reviews. This will help the industry to evaluate products more accurately and consumers will be smarter when making purchase decisions.

Second, compared to luxury and chain hotels, budget hotels gain more from social influence. Exposing customers to prior ratings is beneficial in their situation. The positive effects of social influence do not change for luxury hotels, so they do not need to worry as much about them.

5.3 Research Limitations and Future Research Directions

This dissertation has several limitations that can be addressed. First, even though eight emotions were collected from reviews, only two emotions were included in our model due to strong correlations between emotion dimensions. There is a need to look into other discrete emotions that can be found embedded in reviews, even though joy and anger are the most frequently encountered emotions in review texts (e.g., sadness, fear, surprise).

Furthermore, while our research model considered a number of important factors related to social influence in online reviews, it didn't take into account some details about reviewer characteristics. Future studies can look into the effects of these variables, including the nationality and gender of the reviewers.

Second, the data set was obtained from TripAdvisor (New York hotels) between 2004 and 2011, and it is important to confirm whether these conclusions can be extrapolated to other countries and more recent time periods. Future research may explore a cross-cultural study of the moderators of the social influence effect.

Third, the use of luxury hotels within the same chain was a limitation.
To check the consistency of the results, more luxury hotels can be added. Moreover, future research can include reviewer characteristics to further profile the extracted topics and provide an explanation for satisfaction ratings.

References

Abrate, G., Capriello, A.,& Fraquelli, G. (2011). When quality signals talk: Evidence from the Turin hotel industry. *Tourism Management*, 32(4), 912-921.

Abrate, G., Quinton, S. & Pera, R. (2021). The relationship between price paid and hotel review ratings: Expectancy-disconfirmation or placebo effect?. *Tourism Management*, 85, 104314.

Achar, C., So, J., Agrawal, N., & Duhachek, A. (2016). What we feel and why we buy: the influence of emotions on consumer decision-making. *Current Opinion in Psychology*, *10*, 166-170.

Adamopoulos, P., Ghose, A., & Tuzhilin, A. (2021). Heterogeneous demand effects of recommendation strategies in a mobile application: Evidence from econometric models and machine-learning instruments. *arXiv preprint arXiv:2102.10468*.

Adomavicius, G., Bockstedt, J. C., Curley, S. P., & Zhang, J. (2013). Do recommender systems manipulate consumer preferences? A study of anchoring effects. *Information Systems Research*, *24*(4), 956-975.

Aguiar-Quintana, T., Nguyen, T. H. H., Araujo-Cabrera, Y., & Sanabria-Díaz, J. M. (2021). Do job insecurity, anxiety and depression caused by the COVID-19 pandemic influence hotel employees' self-rated task performance? The moderating role of employee resilience. *International Journal of Hospitality Management*, 94, 102868.

Alzate, M., Arce-Urriza, M., & Cebollada, J. (2022). Mining the text of online consumer reviews to analyze brand image and brand positioning. *Journal of Retailing and Consumer Services*, 67, 102989.

Antonio, N., de Almeida, A. M., Nunes, L., Batista, F., & Ribeiro, R. (2018). Hotel online reviews: creating a multi-source aggregated index. *International Journal of Contemporary Hospitality Management*, *30*(12), 3574-3591.

Ashmore, R. D., Deaux, K., & McLaughlin-Volpe, T. (2004). An organizing framework for collective identity: articulation and significance of multidimensionality. *Psychological bulletin*, *130*(1), 80.

Arnould, E. J., & Price, L. L. (1993). River magic: Extraordinary experience and the extended service encounter. *Journal of Consumer Research*, *20*(1), 24-45.

Baas, M., & Dreu, C., & Nijstad, B.A., (2012). Emotions that associate with uncertainty lead to structured ideation. *Emotion*. 12, 1004-1014.

Babic´Rosario,Ana,B.,FrancescaSotgiu,KristineDeValck,and Tammo H. A. Bijmolt (2016), "The Effect of Electronic Word of Mouth on Sales: A Meta-Analytic Review of Platform, Product, and Metric Factors," *Journal of Marketing Research*, 53 (3), 297-318.

Backaler, J., & Shankman, P. (2018). *Digital influence*. Macmillan.

Baek, J., Choe, Y., & Ok, C. M. (2020). Determinants of hotel guests' service experiences: an examination of differences between lifestyle and traditional hotels. *Journal of Hospitality Marketing & Management*, *29*(1), 88-105.

Bagozzi, R. P., Gopinath, M., & Nyer, P. U. (1999). The role of emotions in marketing. *Journal of the Academy of Marketing Science*, *27*(2), 184-206.

Baka, V. (2016). The becoming of user-generated reviews: Looking at the past to understand the future of managing reputation in the travel sector. *Tourism Management*, 53, 148–162. https://doi.org/10.1016/j.tourman.2015.09.004

Banerjee, S., & Chua, A. Y. (2016). In search of patterns among travellers' hotel ratings in TripAdvisor. *Tourism Management*, 53, 125-131.

Banerjee, S., Bhattacharyya, S., & Bose, I. (2017). Whose online reviews to trust? Understanding reviewer trustworthiness and its impact on business. *Decision Support Systems*, 96, 17-26.

Bao, T., & Chang, T. L. S. (2014). Why Amazon uses both the New York Times Best Seller List and customer reviews: An empirical study of multiplier effects on product sales from multiple earned media. *Decision Support Systems*, *67*, 1-8.

Bigné, J. E., Mattila, A. S., & Andreu, L. (2008). The impact of experiential consumption cognitions and emotions on behavioral intentions. *Journal of Services Marketing,* 22(4), 303-315.

Bastiaansen, M., Lub, X. D., Mitas, O., Jung, T. H., Ascenção, M. P., Han, D. I., ... & Strijbosch, W. (2019). Emotions as core building blocks of an experience. *International Journal of Contemporary Hospitality Management*, 31(2), 651-668.

Basias, N.; Pollalis, Y. Quantitative and qualitative research in business & technology: Justifying a suitable research methodology. Rev. Integr. Busi. Econo. Res. 2018, 7, 91–105.

Bravo, R., Martinez, E., & Pina, J. M. (2019). Effects of service experience on customer responses to a hotel chain. *International Journal of Contemporary Hospitality Management*. 31(1), 389-405.

Browning, V., So, K. K. F., & Sparks, B. (2013). The influence of online reviews on consumers' attributions of service quality and control for service standards in hotels. *Journal of Travel & Tourism Marketing*, 30(1-2), 23-40.

Brown, J., Broderick, A. J., & Lee, N. (2007). Word of mouth communication within online communities: Conceptualizing the online social network. *Journal of interactive marketing*, 21(3), 2-20.

Brehm, J.W. (1966). A theory of psychological reactance. Academic Press.

Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, *49*(2), 192-205.

Berezina, K., Bilgihan, A., Cobanoglu, C., & Okumus, F. (2016). Understanding Satisfied and Dissatisfied Hotel Customers: Text Mining of Online Hotel Reviews. *Journal of Hospitality Marketing & Management*, 25(1), 1–24. https://doi.org/10.1080/19368623.2015.983631

Blei, D. M., Ng, A. Y., & Edu, J. B. (2003). Latent Dirichlet Allocation Michael I. Jordan. *Journal of Machine Learning Research*, 3, 993–1022. https://doi.org/10.5555/944919.944937

Blei, D. M. (2012). Probabilistic topic models. Communications of the ACM, 55(4), 77-84.

Bless, H., Schwarz, N., & Kemmelmeier, M. (1996). Mood and stereotyping: Affective states and the use of general knowledge structures. *European Review of Social Psychology*, *7*(1), 63-93.

Bless, H., Clore, G.L., Schwarz, N., Golisano, V., Rabe, C. & Wölk, M. (1996). Mood and the use of scripts: Does a happy mood really lead to mindlessness. *Journal of Personality and Social Psychology*, 71(4), 665-679.

Bodenhausen, G. V., Sheppard, L. A., & Kramer, G. P. (1994). Negative affect and social judgment: The differential impact of anger and sadness. *European Journal of Social Psychology*, *24*(1), 45-62.

Bodenhausen, G. V., Kramer, G. P., & Süsser, K. (1994). Happiness and stereotypic thinking in social judgment. *Journal of Personality and Social Psychology*, *66*(4), 621-632.

Bodenhausen, G. V., Gabriel, S., & Lineberger, M. (2000). Sadness and susceptibility to judgmental bias: The case of anchoring. *Psychological Science*, *11*(4), 320-323.

Cialdini, R.B. (2009), Influence: Science and Practice, 5th ed., Pearson education, Boston, MA.

Cicognani, S., Figini, P., & Magnani, M. (2016). Social influence bias in online ratings: A field experiment.

Cheng, M., & Jin, X. (2019). What do Airbnb users care about? An analysis of online review comments. *International Journal of Hospitality Management*, 76, 58–70. https://doi.org/10.1016/J.IJHM.2018.04.004

Chen, M. & Farn, C. (2020) Examining the Influence of Emotional Expressions in Online Consumer Reviews on Perceived Helpfulness. *Information Processing and Management*, 57, 102266.

Cheung, C. M., & Thadani, D. R. (2012). The impact of electronic word-of-mouth communication: A literature analysis and integrative model. *Decision support systems*, 54(1), 461-470.

Chevalier, Judith A. and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43 (3), 345-359.

Chintagunta, P. K., Gopinath, S., & Venkataraman, S. (2010). The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing science*, 29(5), 944-957.

Choi, S., Mattila, A. S. (2004). Hotel revenue management and its impact on customers' perceptions of fairness. *Journal of Revenue and Pricing Management*, 2, 303-314.

Crotts, J. C., Mason, P. R., & Davis, B. (2009). Measuring Guest Satisfaction and Competitive Position in the Hospitality and Tourism Industry: An Application of Stance-Shift Analysis to Travel Blog Narratives. *Journal of Travel Research*, 48(2), 139–151. https://doi.org/10.1177/0047287508328795

Dai, H., Chan, C. & Mogilner, C. (2020). People Rely Less on Consumer Reviews for Experiential than Material Purchases. *Journal of Consumer Research*, 46(6), 1052–1075.

Dellarocas, Chrysanthos, Xiaoquan M. Zhang, and Neveen Awad (2007), "Exploring the Value of Online Product Reviews in Fore- casting Sales: The Case of Motion Pictures," *Journal of Interactive Marketing*, 21 (4), 23-45.

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology*, *51*(3), 629-636.

Dev, C., M. S. Morgan, and S. Shoemaker. (1995). A positioning analysis of hotel brands. *Cornell Hotel and Restaurant Administration Quarterly*. 36 (6), 48-55.

Dinçer, M. Z., & Alrawadieh, Z. (2017). Negative Word of Mouse in the Hotel Industry: A Content Analysis of Online Reviews on Luxury Hotels in Jordan. *Journal of Hospitality Marketing & Management*, 26(8), 785–804. https://doi.org/10.1080/19368623.2017.1320258

Ding, K., Choo, W. C., Ng, K. Y., & Ng, S. I. (2020). Employing structural topic modelling to explore perceived service quality attributes in Airbnb accommodation. *International Journal of Hospitality Management*, 91, 102676. https://doi.org/10.1016/J.IJHM.2020.102676

Donato, C., & Miceli, G. (2020). The effect of disgust on preference for structure: Evidence for a double-sided response. *Journal of Consumer Behaviour*, *19*(1), 68-79.

Ekman, P. (1992). An Argument for Basic Emotions. *Cognition & Emotion*, 6 (3-4), 169–200.

Englich, B., & Soder, K. (2009). Moody experts---How mood and expertise influence judgmental anchoring. *Judgment and Decision making*, *4*(1), 41-50.

eMarketer (2019), “What Makes US Gen Z and Millennial Internet Users More Likely to Buy Products/Services that an Influencer Recommends? (% of respondents, Sep 2019),” (November 5), https://chart-na1.emarketer.com/233274/what-makes-us-gen-z-millennial-internet-users-more-likely-buy-productsservices-that-influen cer-recommends-of-respondents-sep-2019.

Epley, N. & Gilovich, T. (2006) The Anchoring-and-Adjustment Heuristic Why the Adjustments Are Insufficient. *Psychological Science*, 17(4), 311-318

Estrada, C. A., Isen, A. M., & Young, M. J. (1997). Positive affect facilitates integration of information and decreases anchoring in reasoning among physicians. *Organizational Behavior and Human Decision Processes*, *72*(1), 117-135.

Felbermayr, A., & Nanopoulos, A. (2016). The role of emotions for the perceived usefulness in online customer reviews. *Journal of Interactive Marketing*, *36*, 60-76.

Filieri, R., Vitari, C.,& Raguseo, E. (2021) Extreme Negative Rating and Review Helpfulness: The Moderating Role of Product Quality. *Journal of Travel Research*, 60(4), 699-717.

Forgas, J. P. (1995). Mood and judgment: the affect infusion model (AIM). *Psychological bulletin*, *117*(1), 39-66.

Forman, C., Ghose, A. & Wiesenfeld, B. (2008). Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets. *Information Systems Research*, 19 (3), 291–313.

Francesco, G., & Roberta, G. (2019). Cross-country analysis of perception and emphasis of hotel attributes. *Tourism Management*, 74, 24-42.

Fromkin, H. L. (1970). Effects of experimentally aroused feelings of undistinctiveness upon valuation of scarce and novel experiences. *Journal of Personality and Social Psychology*, *16*(3), 521-529.

Fromkin, H.L., Snyder, C.R. (1980). The Search for Uniqueness and Valuation of Scarcity. In: Gergen, K.J., Greenberg, M.S., Willis, R.H. (eds) Social Exchange. Springer, Boston, MA.

Gang, R. & Taeho, H. (2019). Examining the relationship between specific negative emotions and the perceived helpfulness of online reviews. *Information Processing and Management,* 56, 1425–1438.

Gao, B., Hu, N., & Bose, I. (2017). Follow the herd or be myself? An analysis of consistency in behavior of reviewers and helpfulness of their reviews. *Decision Support Systems*, *95*, 1-11.

Gao, B., Li, X., Liu, S. & Fang, D. (2018). How power distance affects online hotel ratings: The positive moderating roles of hotel chain and reviewers' travel experience. *Tourism Management*, 65, 176-186.

Gavilan,D., Avello, M., Martinez-Navarro, G. (2018). The influence of online ratings and reviews on hotel booking consideration. *Tourism Management*, 66, 53-61.

Geetha, M., Singha, P., & Sinha, S. (2017). Relationship between customer sentiment and online customer ratings for hotels-An empirical analysis. *Tourism Management*, *61*, 43-54.

Giatsoglou, M., Vozalis, M. G., Diamantaras, K., Vakali, A., Sarigiannidis, G., & Chatzisavvas, K. C. (2017). Sentiment analysis leveraging emotions and word embeddings. *Expert Systems with Applications*, *69*, 214-224.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). Heuristics and biases: The psychology of intuitive judgment. Cambridge university press.

Godes, D., & Mayzlin, D. (2004). Using Online Conversations to Study Word-of-Mouth Communication. Marketing Science, 23(4), 469–631. https://doi.org/10.1287/MKSC.1040.0071

Goldstein, N. J., & Cialdini, R. B. (2009). Normative influences on consumption and conservation behaviors.

Guo, B., & Zhou, S. (2016). Understanding the impact of prior reviews on subsequent reviews: The role of rating volume, variance and reviewer characteristics. *Electronic Commerce Research and Applications*, *20*, 147-158.

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.

Gu, B., & Ye, Q. (2014). First Step in Social Media: Measuring the Influence of Online Management Responses on Customer Satisfaction. *Production and Operations Management*, 23(4), 570–582. https://doi.org/10.1111/POMS.12043

Gu, H., & Ryan, C. (2008). Place attachment, identity and community impacts of tourism—the case of a Beijing hutong. *Tourism Management*, 29(4), 637–647. https://doi.org/10.1016/J.TOURMAN.2007.06.006

Han, S. & Anderson, C.K. (2020). Customer Motivation and Response Bias in Online Reviews, 61(2), *Cornell Hospitality Quarterly*, 61(2), 142-153

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., & Tatham, R. (2006). Multivariate Data Analysis. Uppersaddle River.

Herr, P. M., Kardes, F. R., & Kim, J. (1991). Effects of word-of-mouth and product-attribute information on persuasion: An accessibility-diagnosticity perspective. *Journal of consumer research*, *17*(4), 454-462.

Ho, Y. C., Wu, J., & Tan, Y. (2017). Disconfirmation effect on online rating behavior: A structural model. *Information Systems Research*, *28*(3), 626-642.

Hong, Y., & Pavlou, P. A. (2014). Product fit uncertainty in online markets: Nature, effects, and antecedents. *Information Systems Research*, *25*(2), 328-344.

Hornsey, M. J., Oppes, T., & Svensson, A. (2002). "It's OK if we say it, but you can't": Responses to intergroup and intragroup criticism. *European Journal of Social Psychology*, *32*(3), 293-307.

Hortaçsu, Ali, F. Asís Martínez-Jerez, and Jason Douglas (2009), "The Geography of Trade in Online Transactions: Evidence from eBay and Mercadolibre," American Economic Journal: Microeconomics, 1 (1), 53–74.

Hu, X., Chen, X., & Davison, R. M. (2019). Social support, source credibility, social influence, and impulsive purchase behavior in social commerce. *International Journal of Electronic Commerce*, *23*(3), 297-327.

Hu, X., & Yang, Y. (2021). What makes online reviews helpful in tourism and hospitality? A bare-bones meta-analysis. *Journal of hospitality marketing & management*, *30*(2), 139-158.

Hu, Y, & Li, X. (2011). Context-Dependent Product Evaluations: An Empirical Analysis of Internet Book Reviews. *Journal of Interactive Marketing*, 25, 123-133

Hu, N., Zhang, T., Gao, B., & Bose, I. (2019). What do hotel customers complain about? Text analysis using structural topic model. *Tourism Management*, 72, 417–426. https://doi.org/10.1016/j.tourman.2019.01.002

Hu, N., Koh, N. S., & Reddy, S. K. (2014). Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales. *Decision support systems*, *57*, 42-53.

Huang, B., Ou, Y., & Carley, K. M. (2018). Aspect level sentiment classification with attention-over-attention neural networks. In *Social, Cultural, and Behavioral Modeling: 11th International Conference, SBP-BRiMS 2018, Washington, DC, USA, July 10-13, 2018, Proceedings 11* (pp. 197-206). Springer International Publishing.

Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the international AAAI conference on web and social media. 8(1), 216-225.

Ibrekk, H., & Morgan, M. G. (1987). Graphical communication of uncertain quantities to nontechnical people. Risk analysis, 7(4), 519-529.

Isen, A. M., & *Means*, B. (*1983*). The influence of positive affect on decision-making strategy. *Social Cognition*, 2(1), 18–31

Inbar, Y. & Gilovich, T. (2011) Angry (or Disgusted), but Adjusting? The Effect of Specific Emotions on Adjustment From Self-Generated Anchors, *Social Psychological and Personality Science,* 2(6) 563-569.

Jain, P. K., Pamula, R., & Srivastava, G. (2021). A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Computer science review*, *41*, 100413.

Jani, D., & Han, H. (2011). Investigating the key factors affecting behavioral intentions: Evidence from a full-service restaurant setting. *International Journal of Contemporary Hospitality Management*, 23(7), 1000-1018.

Jani, D., & Han, H. (2015). Influence of environmental stimuli on hotel customer emotional loyalty response: Testing the moderating effect of the big five personality factors. *International Journal of Hospitality Management*, 44, 48-57.

Jiang, H., Liang, J., Wang, H., & Sun, P. (2016). The interplay of emotions, elaboration, and ambivalence on attitude–behavior consistency. *Journal of Consumer Behaviour*, 15(2), 126–135.

Kandampully, J., & Suhartanto, D. (2003). The role of customer satisfaction and image in gaining customer loyalty in the hotel industry. *Journal of Hospitality & Leisure Marketing*, 10(1-2), 3-25.

Kim, J., and Gupta, P. (2012). "Emotional Expressions in Online User Reviews: How They Influence Consumers' Product Evaluations," *Journal of Business Research,* 65(7), 985–992.

Kim, W. G., Lim, H., & Brymer, R. A. (2015). The effectiveness of managing social media on hotel performance. *International Journal of Hospitality Management*, 44, 165–171. https://doi.org/10.1016/j.ijhm.2014.10.014

Kirk, D. (1995). Environmental management in hotels. *International Journal of Contemporary Hospitality Management*, 7(6), 3-8.

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

Köcher, S., & Köcher, S. (2021). The Mode Heuristic in Service Consumers' Interpretations of Online Rating Distributions. *Journal of Service Research*, 24(4), 582-600.

Kuan, K. K., Zhong, Y., & Chau, P. Y. (2014). Informational and normative social influence in group-buying: Evidence from self-reported and EEG data. *Journal of Management Information Systems*, *30*(4), 151-178.

Kwon, H. J., Ban, H. J., Jun, J. K., & Kim, H. S. (2021). Topic modeling and sentiment analysis of online review for airlines. *Information*, 12(2), 78.

Laros, F.J.M. & Steenkamp, J.E.M. (2005). Emotions in consumer behavior: a hierarchical approach. *Journal of Business Research*, 58, 1437-1445.

Lee, K., & Yu, C. (2018). Assessment of airport service quality: A complementary approach to measure perceived service quality based on Google reviews. *Journal of Air Transport Management*, 71, 28–44. https://doi.org/10.1016/J.JAIRTRAMAN.2018.05.004

Lee, Y. J., Hosanagar, K., & Tan, Y. (2015). Do I follow my friends or the crowd? Information cascades in online movie ratings. *Management Science*, *61*(9), 2241-2258.

Lee, M., Lee, S.A. and Koh, Y. (2019), "Multisensory experience for enhancing hotel guest experience", *International Journal of Contemporary Hospitality Management*, 31(11), 4313-4337.

Lelieveld, G. & Hendriks, H. (2021) The interpersonal effects of distinct emotions in online reviews, *Cognition and Emotion*, 35(7), 1257-1280.

Lerner, J.S. & Keltner, D. (2000). Beyond valence: Toward a model of emotion-specific influences on judgement and choice, *Cognition & Emotion*, 14(4), 473-493

Lerner, J. S., & Keltner, D. (2001). Fear, anger, and risk. *Journal of Personality and Social Psychology*, *81*(1), 146-159.

Lerner, J. S., & Tiedens, L. Z. (2006). Portrait of the angry decision maker: How appraisal tendencies shape anger's influence on cognition. *Journal of Behavioral Decision Making*, *19*(2), 115-137.

Lerner, J. S., Goldberg, J. H., & Tetlock, P. E. (1998). Sober second thought: The effects of accountability, anger, and authoritarianism on attributions of responsibility. *Personality and Social Psychology Bulletin*, 6, 563-574.

Li, X., & Hitt, L. M. (2008). Self-selection and information role of online product reviews. *Information Systems Research*, *19*(4), 456-474.

Li, H., Meng, F., Jeong, M., & Zhang, Z. (2020). To follow others or be yourself? Social influence in online restaurant reviews. *International Journal of Contemporary Hospitality Management*, 32(3), 1067-1087

Li, H., Zhang, Z., Meng, F., & Zhang, Z. (2019). "When you write review" matters: The interactive effect of prior online reviews and review temporal distance on consumers' restaurant evaluation. *International Journal of Contemporary Hospitality Management*, 31(3),1273-1291

Li, G., Law, R., Vu, H. Q., Rong, J., & Zhao, X. (Roy). (2015). Identifying emerging hotel preferences using Emerging Pattern Mining technique. *Tourism Management*, 46, 311–321. https://doi.org/10.1016/J.TOURMAN.2014.06.015

Li, H., Ye, Q., & Law, R. (2013). Determinants of Customer Satisfaction in the Hotel Industry: An Application of Online Review Analysis. *Asia Pacific Journal of Tourism Research*, 18(7), 784–802. https://doi.org/10.1080/10941665.2012.708351

Liu, Yong (2006), "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," *Journal of Marketing*, 70 (3), 74-89.

Lyberg, L.E.; Weisberg, H.F.; Wolf, C.; Joye, D.; Smith, T.; Fu, Y.-C. Total Survey Error: A Paradigm for Survey Methodology. In The SAGE Handbook of Survey Methodology; SAGE Publications Pvt Ltd: Thousand Oaks, CA, USA, 2016; pp. 27–40.

Ma, Liye, Ramayya Krishnan, and Alan L. Montgomery (2014), "Latent Homophily or Social Influence? An Empirical Analysis of Purchase Within a Social Network," Management Science, 61 (2), 454–73.

Ma, X., Khansa L., K.,Yun, D., & Sung K. (2014). Impact of Prior Reviews on the Subsequent Review Process in Reputation Systems. *Journal of Management Information Systems,* 30(3), 279–310

Mackie, D. M., & Worth, L. T. (1989). Processing deficits and the mediation of positive affect in persuasion. *Journal of Personality and Social Psychology*, *57*(1), 27.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., & McClosky, D. (2014, June). The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations* (pp. 55-60).

Mariani, M.M. & Borghi, M. (2018). Effects of the Booking.com rating system: Bringing hotel class into the picture. *Tourism Management*, 66(June), 47-52.

Mariani, M. (2019). Big data and analytics in tourism and hospitality: a perspective article. *Tourism Review*.

Martin-Fuentes, E. (2016). Are guests of the same opinion as the hotel star-rate classification system?. *Journal of Hospitality and Tourism Management*, 29, 126-134

Martin, L., Sintsova, V., & Pu, P. (2014). Are influential writers more objective? An analysis of emotionality in review comments. In *Proceedings of the 23rd International Conference on World Wide Web,April,* 799-804).

Mayzlin, D., Dover, Y, & Chevalier, J. (2014). Promotional Reviews: An Empirical Investigation of Online Review Manipulation. *American Economic Review*, 104(8), 2421–2455

Mellinas, J. P., Nicolau, J. L., & Park, S. (2019). Inconsistent behavior in online consumer reviews: The effects of hotel attribute ratings on location. *Tourism Management*, 71, 421–427. https://doi.org/10.1016/j.tourman.2018.10.034

Meyners, Jannik, Christian Barrot, Jan U. Becker, and Jacob Goldenberg (2017), "The Role of Mere Closeness: How Geographic Proximity Affects Social Influence," Journal of Marketing, 81 (5), 49–66.

Moe, W. W., & Trusov, M. (2011). The value of social dynamics in online product ratings forums. *Journal of Marketing Research*, 48(3), 444-456.

Moe, W. W., & Schweidel, D. A. (2012). Online product opinions: Incidence, evaluation, and evolution. *Marketing Science*, *31*(3), 372-386.

Mohammad, S. & Turney, P. (2010). Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, 26–34, Los Angeles, CA.

Mohammad, S. M., & Turney, P. D. (2013). CROWDSOURCING A WORD–EMOTION ASSOCIATION LEXICON. Computational Intelligence, 29(3), 436–465. https://doi.org/10.1111/J.1467-8640.2012.00460.X

Mok, Diana, Barry Wellman, and Juan Carrasco (2010), "Does Distance Matter in the Age of the Internet?" Urban Studies (Edinburgh, Scotland), 47 (13), 2747–83.

Moons, W. G., & Mackie, D. M. (2007). Thinking straight while seeing red: The influence of anger on information processing. *Personality and Social Psychology Bulletin*, *33*(5), 706-720.

Murgado-Armenteros, E. M., Torres-Ruiz, F. J., & Vega-Zamora, M. (2012). Differences between online and face to face focus groups, viewed through two approaches. *Journal of theoretical and applied electronic commerce research*, 7(2), 73-86.

Muchnik, L., Aral, S., & Taylor, S. J. (2013). Social influence bias: A randomized experiment. *Science*, *341*(6146), 647-651.

Namasivayam, K., Enz, C. A., & Siguaw, J. A. (2000). How wired are we? The selection and use of new technology in US Hotels. *The Cornell Hotel and Restaurant Administration Quarterly*, 41(6), 40-48.

O'Connor, P. (2008). User-Generated Content and Travel: A Case Study on Tripadvisor.Com. In: O'Connor, P., Höpken, W., Gretzel, U. (eds) Information and Communication Technologies in Tourism 2008. Springer, Vienna.

O'Connor, P. (2010). Managing a hotel's image on TripAdvisor. *Journal of Hospitality Marketing & Management*, *19*(7), 754-772.

O'Neill, J.W. & Carlback, M. (2011). Do brands matter? A comparison of branded and independent hotels' performance during a full economic cycle, *International Journal of Hospitality Management*, 30 (3), 515-521.

Olorunniwo, F., Hsu, M. K., & Udo, G. J. (2006). Service quality, customer satisfaction, and behavioral intentions in the service factory. *Journal of services marketing*, *20*(1), 59

Özdağoğlu, G., Kapucugil-İkiz, A., & Çelik, A. F. (2018). Topic modelling-based decision framework for analysing digital voice of the customer. *Total Quality Management & Business Excellence*, 29(13-14), 1545-1562.

Palese, B., & Usai, A. (2018). The relative importance of service quality dimensions in E-commerce experiences. *International Journal of Information Management*, 40, 132–140. https://doi.org/10.1016/J.IJINFOMGT.2018.02.001

Park, C., & Lee, T. M. (2009). Information direction, website reputation and eWOM effect: A moderating role of product type. Journal of Business research, 62(1), 61-67.

Park, D. H., Lee, J., & Han, I. (2007). The effect of on-line consumer reviews on consumer purchasing intention: The moderating role of involvement. International journal of electronic commerce, 11(4), 125-148.

Pentina, I., Bailey, A. A., & Zhang, L. (2018). Exploring effects of source similarity, message valence, and receiver regulatory focus on yelp review persuasiveness and purchase intentions. Journal of Marketing Communications, 24(2), 125-145.

Parasuraman, A., Zeithaml, V. A., & Berry, L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *1988*, *64*(1), 12-40.

Phillips, P., Zigan, K., Santos Silva, M. M., & Schegg, R. (2015). The interactive effects of online reviews on the determinants of Swiss hotel performance: A neural network analysis. *Tourism Management*, 50, 130–141. https://doi.org/10.1016/j.tourman.2015.01.028

Poncheri, R.M., Lindberg, J.T., Thompson, L.F., & Surface, Eric A. (2008). A Comment on Employee Surveys: Negativity Bias in Open-Ended Responses. *Organisational Research Methods*, 11(3), 614-630.

Prebensen, N.K. and Rosengren, S. (2016). Experience value as a function of hedonic and utilitarian dominant services. *International Journal of Contemporary Hospitality Management*, 28(1), 113-135.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In *Theories of emotion* (pp. 3-33). Academic press.

Puccinelli, N. M., Goodstein, R. C., Grewal, D., Price, R., Raghubir, P., & Stewart, D. (2009). Customer experience management in retailing: understanding the buying process. *Journal of retailing*, *85*(1), 15-30.

Putranto, Y., Sartono, B., & Djuraidah, A. (2021). Topic modelling and hotel rating prediction based on customer review in Indonesia. *International Journal of Management and Decision Making*, 20(3), 282–307. https://doi.org/10.1504/IJMDM.2021.116028

Quinn, Annalisa (2018), "Thought Leaders," The New York Times Magazine (November 20), https://www.nytimes.com/2018/11/20/magazine/everyone-wants-to-influence-you.html.

Radojevic,T., Stanisic, N. & Stanic, N. (2015). Ensuring positive feedback: Factors that influence customer satisfaction in the contemporary hospitality industry. *Tourism Management*, 51, 13-21.

Ren, G. & Hong, T. (2019). Examining the relationship between specific negative emotions and the perceived helpfulness of online reviews. *Information Processing & Management,* 56(4), 1425-1438.

Rhee, H.T., & Yang, S. (2015), Does hotel attribute importance differ by hotel? Focusing on hotel star-classifications and customers' overall ratings *Computers in Human Behavior*, 50, 576–587.

Reisenzein, R. (1994). Pleasure-arousal theory and the intensity of emotions. *Journal of Personality and Social Psychology*, *67*(3), 525-539.

Ric, F. (2004). Effects of the activation of affective information on stereotyping: When sadness increases stereotype use. *Personality and Social Psychology Bulletin*, 30, 1310-1321.

Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2016). A Model of Text for Experimentation in the Social Sciences. *Journal of the American Statistical Association*, 111(515), 988–1003. https://doi.org/10.1080/01621459.2016.1141684

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, 110, 145-172.

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of Personality and Social Psychology*, *76*(5), 805.

Russell, J. A., & Mehrabian, A. (1974). Distinguishing anger and anxiety in terms of emotional response factors. *Journal of Consulting and Clinical Psychology*, *42*(1), 79-83.

Schlosser, A. E. (2005). Posting versus lurking: Communicating in a multiple audience context. *Journal of Consumer Research*, 32(2), 260-265.

Schlosser, A. E. (2011). Can including pros and cons increase the helpfulness and persuasiveness of online reviews? The interactive effects of ratings and arguments. *Journal of Consumer Psychology, 21(3), 226-239.*

Schwarz, N. (1990). Feeling as information: Informational and motivational functions of affective states. In E. T. Higgins & R. M. Sorrentino (Eds.), Handbook of motivation and cognition (Vol. 2., pp. 527-561). New York: Guilford.

Schwarz, N. (1998). Accessible content and accessibility experiences: The interplay of declarative and experiential information in judgment. *Personality and Social Psychology Review*, *2*(2), 87-99.

Sherif, M. (1936). The psychology of social norms. Harper, Oxford, England.

Singh, J. P., Irani, S., Rana, N. P., Dwivedi, Y. K., Saumya, S., & Roy, P. K. (2017). Predicting the “helpfulness” of online consumer reviews. *Journal of Business Research*, 70, 346-355.

Sridhar, S. and Srinivasan, R. (2012), “Social influence effects in online product ratings”, *Journal of Marketing*, 76(5), 70-88

Stringam, B. B., & Gerdes, J. (2010). An Analysis of Word-of-Mouse Ratings and Guest Comments of Online Hotel Distribution Sites. *Journal of Hospitality Marketing & Management* , 19(7), 773–796. https://doi.org/10.1080/19368623.2010.508009

Tajfel, H. (1974). Social identity and intergroup behaviour. *Social science information*, *13*(2), 65-93.

Takhteyev, Yuri, Anatoliy Gruzd, and Barry Wellman (2012), "Geography of Twitter Networks," Social Networks, 34 (1), 73–81.

Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: the mobilization-minimization hypothesis. *Psychological bulletin*, *110*(1), 67-85.

Todri, V., Adamopoulos, P., & Andrews, M. (2022). Is distance really dead in the online world? The moderating role of geographical distance on the effectiveness of electronic word of mouth. *Journal of Marketing*, *86*(4), 118-140.

Tiedens, L.Z. & Lincon, S. (2001). Judgment Under Emotional Certainty and Uncertainty: the Effects of Specific Emotions on Information Processing. *Attitudes and Social Cognition*. 18, 973-988.

Turner, J. C., & Oakes, P. J. (1986). The significance of the social identity concept for social psychology with reference to individualism, interactionism and social influence. *British Journal of Social Psychology*, *25*(3), 237-252.

Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science* 185, 1124–1131.

Ullah, R., Zeb, A., & Kim, W. (2015). The impact of emotions on the helpfulness of movie reviews. *Journal of Applied Research and Technology*, *13*(3), 359-363.

Van Kleef, G. A. (2014). Understanding the positive and negative effects of emotional expressions in organizations: EASI does it. *Human Relations*, *67*(9), 1145-1164.

Vermeulen, I.E. & Seegers, D. (2009). Tried and tested: the impact of online hotel reviews on consumer consideration. *Tourism Management*, 30(1), 123–127.

Viglia, G., Minazzi, R., Buhalis, D., (2016). The influence of e-word-of-mouth on hotel occupancy rate. *International Journal of Contemporary Hospitality Management*, 28(9),2035-2051.

Wang, C., Zhang, X., & Hann, I. H. (2018). Socially nudged: A quasi-experimental study of friends' social influence in online product ratings. *Information Systems Research*, *29*(3), 641-655.

Watson, L., & Spence, M. T. (2007). Causes and consequences of emotions on consumer behaviour: A review and integrative cognitive appraisal theory. *European Journal of Marketing*, 41(5/6), 487–511.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin, 98*(2), 219–235.

Weber, E., & Kirsner, B. (1997). Reasons for rank-dependent utility evaluation. *Journal of Risk and Uncertainty*, 14(1), 41-61.

Wegener, D. T., & Petty, R. E. (1994). Mood management across affective states: The hedonic contingency hypothesis. *Journal of Personality and Social Psychology*, 66, 1034-1048.

Westbrook, R. A., & Oliver, R. L. (1991). The dimensionality of consumption emotion patterns and consumer satisfaction. *Journal of Consumer Research*, *18*(1), 84-91.

Wetzer, I. M., Zeelenberg, M., & Pieters, R. (2007). "Never eat in that restaurant, I did!": Exploring why people engage in negative word-of-mouth communication. *Psychology & Marketing*, *24*(8), 661-680.

Wang,X., Tang,L.,& Kim, E. (2019) More than words: Do emotional content and linguistic style matching matter on restaurant review helpfulness? *International*

Journal of Hospitality Management, 77, 438-447

Wang, C., Zhang, X., & Hann, I. H. (2018). Socially nudged: A quasi-experimental study of friends' social influence in online product ratings. *Information Systems Research*, *29*(3), 641-655.

Wu, L., Hoi, S. C. H., & Yu, N. (2010a). Semantics-Preserving Bag-of-Words Models and Applications. IEEE Transactions on Image Processing, 19(7), 1908–1920. https://doi.org/10.1109/tip.2010.2045169

Wu, L., Hoi, S. C. H., & Yu, N. (2010b). Semantics-preserving bag-of-words models and applications. IEEE Transactions on Image Processing, 19(7), 1908–1920. https://doi.org/10.1109/TIP.2010.2045169

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

Xia, L. (2023). The impacts of geographic and social influences on review helpfulness perceptions: A social contagion perspective. *Tourism Management*, *95*, 104687.

Xie, K., & Lee, Y. J. (2015). Social media and brand purchase: Quantifying the effects of exposures to earned and owned social media activities in a two-stage decision making model. *Journal of Management Information Systems*, *32*(2), 204-238.

Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*. 58, 51-65.

Xu, X. (2019). Examining The Relevance Of Online Customer Textual Reviews On Hotels' Product And Service Attributes. *Journal of Hospitality & Tourism Research*, 43(1), 141–163.

Xu, X., Wang, X., Li, Y., & Haghighi, M. (2017). Business intelligence in online customer textual reviews: Understanding consumer perceptions and influential factors. *International Journal of Information Management*, 37(6), 673–683. https://doi.org/10.1016/J.IJINFOMGT.2017.06.004

Xue, F., Dong, L., Gao, B., Yu, Z. and Taras, V. (2020), "Understanding the relationships between distances and herd behavior in online reviews: the moderating effects of hospitality experience", *International Journal of Contemporary Hospitality Management*, 32(10), 3295-3314.

Yacouel, N., & Fleischer, A. (2012). The role of cybermediaries in reputation building and price premiums in the online hotel market. *Journal of Travel Research*, *51*(2), 219-226.

Yan, Q., Zhou, S. & Wu,S. (2018). The influences of tourists' emotions on the selection of electronic word of mouth platforms. *Tourism Management*, 66, 348-363.

Yap, A.J., & Tong, E.M.W. (2009). The Appraisal Rebound Effect: Cognitive Appraisals on the Rebound. *Personality and Social Psychological Bulletin*. 35, 1208 –1219.

Ye, Q., Law, R., & Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management*, *28*(1), 180-182.

Yin, D., Bond, S. D., & Zhang, H. (2014). Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. *MIS Quarterly*, *38*(2), 539-560.

Yin D., Bond S.D. & Zhang H. (2021). Anger in Consumer Reviews: Unhelpful but Persuasive? *MIS Quarterly*, 45(3), 1059-1086.

Yi, S., Li, X., & Jai, T. M. (Catherine). (2016). Hotel guests' perception of best green practices: A content analysis of online reviews: *Tourism and Hospitality Research*, 18(2), 191–202. https://doi.org/10.1177/1467358416637251

Yoo, K. H., & Gretzel, U. (2008). What motivates consumers to write online travel reviews?. *Information Technology & Tourism*, *10*(4), 283-295.

Zeithaml, V.A. 1988 Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence, *Journal of Marketing*,52(3), 2-22

Zhang, Y., Jin, R., & Zhou, Z. H. (2010). Understanding bag-of-words model: a statistical framework. *International Journal of Machine Learning and Cybernetics*, *1*(1), 43-52.

Zhang, Z., Zhang, Z. & Yang, Y. (2016). The power of expert identity: How website-recognized expert reviews influence travelers' online rating behavior, *Tourism Management*, 55, 15-24.

Zhang, J. (2019). Listening to the Consumer: Exploring Review Topics on Airbnb and Their Impact on Listing Performance. *Journal of Marketing Theory and Practice* , 27(4), 371–389. https://doi.org/10.1080/10696679.2019.1644953

Zhang, Y., Jin, R., & Zhou, Z. H. (2010a). Understanding bag-of-words model: a statistical framework. *International Journal of Machine Learning and Cybernetics 2010* 1:1, 1(1), 43–52. https://doi.org/10.1007/S13042-010-0001-0

Zhang, Y., Jin, R., & Zhou, Z.-H. (2010b). Understanding bag-of-words model: a statistical framework. *International Journal of Machine Learning and Cybernetics*, 1(1–4), 43–52. https://doi.org/10.1007/s13042-010-0001-0

Zhao, Y., Xu, X., & Wang, M. (2019). Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews. *International Journal of Hospitality Management*, *76*, 111-121.

Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing*, *74*(2), 133-148.