How much is a scathing review worth?

Karri Hellström

Department of Marketing

Hanken School of Economics

Helsinki

2023

HANKEN SCHOOL OF ECONOMICS

Department: Marketing	Type of work: Master's thesis
Author and student number: **Karri Hellström - 162620**	**Date: 30.05.2023**
Title of Paper: ***How much is a scathing review worth?***	

Abstract:

This paper investigates the impact of customer reviews on hotel prices. More specifically, it aims to verify and quantify the impact of the sentiment expressed in the reviews. This is done by using a novel measure, sentiment intensity, made possible by recent advances in computational linguistics, to examine how the written content acts as a price determinant. After collecting online data about 36 hotels in Helsinki, revenue management is combined with hedonic price theory, to find that the sentiment of the review content matters more than the review ratings, is moderated by time, but is not moderated by high demand.

Keywords: eWOM, online review, hospitality, revenue management, dynamic pricing, hedonic price theory, sentiment analysis

CONTENTS

APPENDICES

TABLES

FIGURES

1 INTRODUCTION

> "If the primary aim of revenue management is selling the right product/service to the right customer, at the right time, for the right price, in order to generate revenue from perishable capacity, then an understanding of consumers and their behaviour is critical to its effective development and implementation" (Yeoman, I., 2016, p. 193)

1.1 The research problem

The decision to book a hotel can be motivated by, among other things, location, brand, facilities, price, service quality, star rating and reviews by past guests (Cantallops & Salvi, 2014; Nieto-Garcia, Munoz-Gallego & Gonzalez-Benito, 2017). Technology has profoundly changed how customers make these decisions. The way customers plan and purchase hotel rooms has changed immensely in the last 20 years (Huiyue, Peihan & Haiwen, 2022; Rita et al., 2022). Disruptive changes have had a fundamental shift in customer behaviour, marketing strategies, distribution and business intelligence (Boccali et al., 2022; Roy et al., 2022). The change caused by digitalization has been even more profound in hospitality than in other businesses (Liang et al., 2022; Majumder, Gupta & Paul, 2022). Potential customers increasingly utilize experiences shared by others to evaluate the reputation of a hotel (Illescas-Manzano et al., 2023; Xu et al., 2019).

The evolution of technology and customers' information seeking behaviour has led to increased market transparency, allowing customers to compare and evaluate options of intangible, subjective, and thus risky products (De Olivera Santos, 2016; Mukhopadhyay, Pandey & Rishi, 2023). In service industries, like hospitality, the experiential features are difficult to evaluate prior to consumption (Kim & Han, 2022; Mehraliyev, Chan & Kirilenko, 2022;), so reviews matter more for services than they do for tangible goods (Yang et al., 2018). Experiential services create a need for customers to search for more information (Gretzel et al., 2020). Reviews are a trusted source for this information for customers, more so than company-produced information nowadays (Abdullah et al., 2022; El-Manstrly, Ali & Line, 2021).

Reviews are a form of electronic word-of-mouth (eWOM), informal communications among non-commercial actors about a service or a good, and one of the most used sources of information used by customers today to facilitate decision-making (Majumder et al., 2022; Mukhopadhyay, Pandey & Rishi, 2023). 90 % of customers read reviews before making a decision to book a hotel (Jiang et al., 2021), Kaemingk (2022) claims that 93% of all online customers read reviews before buying. Currently being the most used way to receive customer feedback, eWOM has transformed how hotels both market their products and run their daily operations (Zhang et al., 2022). Tripadvisor, a third-party review site dedicated to publishing customer reviews, currently boasts having more than one billion reviews on eight million businesses (Tripadvisor, 2023).

The research on eWOM has focused on two main categories: impact it has on customers and companies (Cantallops & Salvi, 2014; Sann et al., 2021). On the customer side, topics like the impact on trust, purchase intention, engagement and satisfaction have been researched (Mukhopadhyay et al., 2023). On the company side, positive reviews have been found to have a positive impact on company performance (Illescas-Manzano et al., 2023; Roy, 2023). For companies, reviews are a way to understand the needs and expectations of customers (Nie et al., 2020; Xu et al., 2019). Analysing reviews allows practitioners to cost efficiently evaluate customers' experiences and value perceptions in real-time, and act accordingly (Boccali et al., 2022; Cheng & Jin, 2019; Neirotti, Raguseo & Paolucci, 2016). For practitioners, better online reviews also allow a better strategic positioning of the hotel, allowing the usage of a better image to acquire a higher price-point (Sánchez-Pérez, Illescas-Manzano & Martínez-Puertas, 2019; Tran, 2020). Positive eWOM improves demand (Binesh, Belarmino & Raab, 2021), average price (Hu, Yang & Park, 2019), sales (Roy et al. 2022; Ye et al., 2011) and occupancy (Phillips et al., 2017). EWOM is an area of importance for researchers too, due to its large impact on customer behaviour, company performance and market dynamics (Roy et al., 2022).

The eWOM impact on company performance has been studied through quantitative measures of review attributes like amount reviews posted, how positive or negative the ratings received are and how much there is variance in these ratings (Zhang et al., 2019). Performance has been quantified through sales and occupancy rate (Mariani & Visani, 2019). The rising impact of online travel agencies and online reviews has created a rapidly shifting landscape, where reacting through pricing is ever more important (Noone, 2016; Yeoman, 2016). While review the relationship between review and performance have been extensively studied, the relationship between price and reviews has not (Abdullah et al. 2022). The research focusing on review impact and companies has mostly studied how eWOM shapes performance, treating price as a factor impacting eWOM, an input to a review, rather than the price being impacted by eWOM (Cantallops & Salvi, 2014; Sann et al., 2021; Yang, Park & Hu, 2018). Price is seen as a component impacting satisfaction, but the impact of said satisfaction on pricing is an under-researched area (Xu, 2019).

Price acts as a signal for quality, guides purchase intention and expectations (Abrate & Viglia, 2016; Abrate, Quinton & Pera, 2021). Price is also how the revenue enters the hotel (Sanchez-Lozano, Pereira & Chávez-Miranda, 2021). It is a crucial element of the marketing mix, as it is the only one that is a revenue producer, not a cost (Arora & Mathur, 2020; Dhurkari, 2023). It includes information about both company and customer behaviour, unlike the other marketing mix elements (Illescas-Manzano et al., 2023). For hotels, pricing heavily drives performance (Vives, Jacob & Aguilo, 2019). In a competitive environment, like it is for hotels, prices are a flexible, easy and fast way to adjust to changing circumstances (Vives & Jacob, 2023). It is one of the tools of revenue management used by hotels to maximize revenue (Vives et al., 2019). Perishable product and limited inventory available, combined with a highly fluctuating demand, creates a fertile ground for using revenue management (RM) (Gibbs, Guttentag & Gretzel, 2018; Vives, Jacob & Aguilo, 2019). Further exasperated by low margins and high fixed costs, RM is utilized to manage the limited inventory and dynamic prices (Vives & Jacob, 2023). Digitalization, automation and the evolution of online travel agencies has had a tremendous impact on pricing strategies, making dynamic, time dependant prices the common practice nowadays (Petricek, Chalupa & Chadt, 2020; Vives et al., 2019). Dynamic pricing captures better the needs of the hotels – to maximize revenue, and the needs of the customers – to acquire a limited inventory product at a specific time (Petricek et al., 2020). Current research in RM focuses on dynamic online pricing, where customer understanding is vital for effective strategies (Petricek, Chalupa & Melas, 2021; Sánchez-Pérez, Illescas-Manzano & Martínez-Puertas, 2019).

Price is a crucial success factor – it being under-researched is a clear gap in the research pertaining to eWOM. While dynamic pricing is well researched in hospitality, there is still work to be done on studying the interplay between pricing and eWOM (Han & Bai, 2022). In pricing literature hotel price determinants are studied through hedonic price theory, which studies the contribution of each individual component, that together make up a product, to its price (Abrate & Viglia, 2016). This body of literature uses eWOM rating as a factor to quantify how much it contributes to the price (see e.g., Abrate & Viglia, 2016; Abrate, Nicolau & Viglia, 2019; Sánchez-Lozano et al., 2021 and Soler et al., 2019).

The usage of ratings only as a price determinant is problematic for two reasons. Firstly, sentiment and content drive sales (Liu et al., 2022; Majumder et al., 2022). The ratings are an aggregated overall score given online to a hotel, of the reviews given to a property (Zhang et al., 2023). These ratings function as a heuristic used by customers in their decision-making, and are used prior to reading the actual unique reviews, to narrow the choices (Gursoy 2019). The emotional intensity expressed in the reviews, positive or negative, is an important factor to consider in the pricing context too (Nie et al., 2020; Zhang et al., 2023). In hospitality, the variable costs are low compared to fixed costs, so a higher price translates almost directly to better the bottom-line results (Leoni & Nilsson, 2021; Matsuoka, 2022). There is evidence that rich content (text and pictures) is a better predictor of sales than numerical ratings (Zhang et al., 2019). While Abrate & Viglia (2016) find empirical support to the hypothesis of a positive impact of online customer's ratings on prices, the content behind these ratings remains unresearched.

The valence of the ratings – how positive or negative the reviews are perceived – has been found to be a significant impactor on customer behaviour, and Yen & Tang (2019) call for more research on review content, as do Qiao et al. (2022). The impact of ratings has been extensively studied, but the study of review content is a relatively new field (Kim & Han, 2022; Roy, 2023; Tran, 2020). The textual reviews include rich information not included in the ratings (Zhang et al., 2019; Qiao et al., 2022), providing more detailed information, lost in the aggregation of the ratings (Mariani, 2020; Nieto-Garcia et al., 2019). This gap in hospitality research has been improved by recent advances in natural language processing, but is still a novel approach and a new way to incorporate the detailed data from the textual reviews into research. It is now possible to incorporate the sentiment of review texts and their intensity into studies about how eWOM impacts performance and pricing (Kwon, 2023). During the literature review, no research was found of the impact of the textual content on price, only that of the ratings. In recent research the sentiment of the review content has been used to find attributes that contribute to hotel performance (Tang & Kim, 2022), forecasting (Chen et al., 2021) and satisfaction (Xu & Zhao, 2022), among many other things. Incorporating new measures of eWOM into research also makes the body of research more generalizable (Sánchez-Pérez, Illescas-Manzano & Martínez-Puertas, 2019).

Second problem with using ratings as a measure is, as customers only read recent reviews (Huiyue et al., 2022), the usage of the overall rating is not an accurate depiction of the review impact from a time perspective. The impact of review recency has been studied from the customer perspective (Liang et al., 2022; Zhang et al., 2023), but has not been incorporated in studies on review impact on company performance, or prices. The element of time and the impact of ratings on price is also a seldom researched area, only a handful of studies were found where the prices were observed as they appear in the industry, accounting for their dynamic nature (see e.g., Abrate, Nicolau & Viglia, 2019; Abrate & Viglia, 2016; Soler et al., 2019). No research was found of sentiment impact on price. This paper aims to address this gap by examining hotel online price determinants and answer the research question:

How does review sentiment impact hotel dynamic prices?

1.2 The conceptual framework of the thesis

Hotels can be classified into three groups: budget, midscale and luxury (Qiao et al., 2022). These classes differentiate from each other in the tangible aspects, like room and bed, as well as intangible aspects, like services and attention from the staff (Roy, 2023). A hotel rooms price can be thought to be determined by six main factors: site (e.g., location, climate, beach, attractions), hotel type (e.g., size, age, number of housekeeping staff), quality signals (e.g., star rating, online rating, chain and brand affiliation), services and amenities (e.g., breakfast, Wi-Fi), property characteristics (e.g., pool, fitness centre) and external market factors (e.g., competition, season) (Alderighi et al., 2022; Bigne, Nicolau & William, 2021; Yang & Leung, 2018).

A hotel room's price can also be thought to be impacted by internal and external factors (Illescas-Manzano et al., 2023; Latinopoulos, 2018). Internal factors, like the service offering, are ones the hotel can influence, external factors, like the online reputation, location and competition, they cannot (Latinopoulos, 2018). These factors act as inputs to a revenue management systems, which calculate the price provided to the customers (Ivanov, 2014; Wamsler, Natter & Algesheimer, 2022). The initial conceptual framework of this paper is depicted in Figure 1 below.

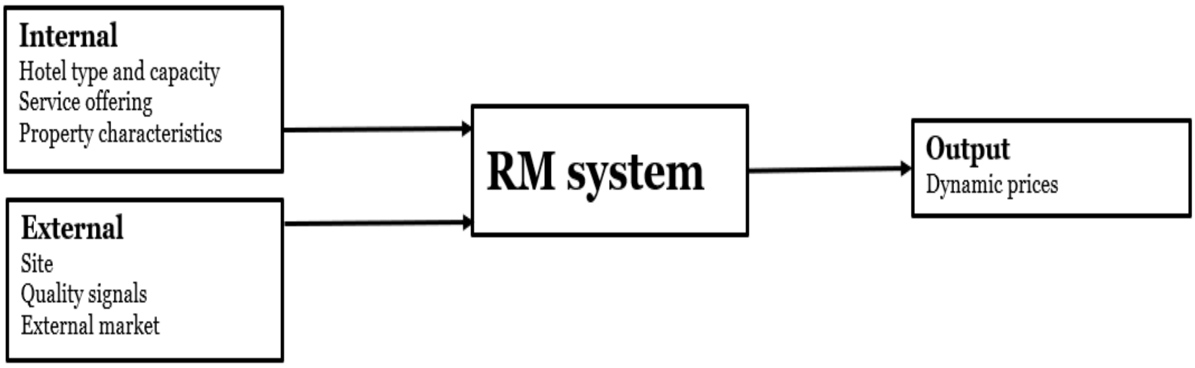

Figure 1: Conceptual framework

1.3 Delimitations

The *industry delimitation:* reviews are readily available online for a variety of products and services, the study focuses on hotels. The reason for this choice is that the dynamic pricing used for pricing hotel rooms based on their demand at certain point in time ensures that shifts in demand are reflected in the price, unlike other consumer goods and services, which have fixed prices for longer periods.

The *geographical delimitation:* is the choice to focus on the Helsinki hotel market. While the methods and results of the study should be easily applicable to other markets, out of lacking previous research, convenience and curiosity the research focuses on the aforementioned market.

The *time delimitation:* all the data for the study was gathered during June and August 2017. Seasonality affects all tourism related industries, but hopefully the methods and their integrity will negate the effects there of. The study attempts to standardize all data used so that the results and methodologies can be utilized throughout the year, and for all markets where online reviews are available.

The *observation delimitation:* only the actual behaviour of the consumers will be observed and analysed, so no internal factors from the customers' mind-set will be studied, only the result of how they have perceived the hotels and their ability to match their offering to the consumer needs. This choice simplifies the study of such a complex phenomenon as reviews, as the causalities or significances of separate different factors need not be accounted for. Only the actual behavioural result - the resulting price – is studied, not all the factors that cause the poor reviews, nor all the factors that, for example, motivate people to post reviews. The researcher has chosen one perspective, the language used in the reviews, and this is used as a summary measure for all other review related aspects that have been identified as meaningful factors of reviews, which impact the company. For example, number of reviews, who wrote the review and why etc.

1.4 Thesis structure

The remainder of the paper is organized as follows. Chapter two contains a literature review and framework expansion with hypothesis formulation. The literature review begins with a short introduction on revenue management, a part of which is dynamic pricing, which is then elaborated. Online travel agencies, which act as a conduit to eWOM and facilitate dynamic pricing are then presented. Furthermore, eWOM, its components, and its impact on companies and customers are then introduced. And finally, hedonic price theory is explained, as it is used to examine the research problem. Chapter three outlines the methods used for the research. Data collection method is outlined thoroughly, the method used to analyse unstructured data is explained, and the variables chosen are introduced. Chapter four presents the results of seven models developed for testing different hypotheses. Chapter five provides an analysis of the models, and finally chapter six contains the research limitations and directions for further research.

2 LITERATURE REVIEW

2.1 Revenue management

> "The process of allocating the right type of capacity to the right kind of customer at the right price so as to maximize revenue or yield" (Kimes, 1989, p. 15)

In hospitality variable costs are low compared to fixed costs, so maximizing revenue can be thought to be equivalent to maximizing profit, and a simple cost-plus pricing approach will not work (Leoni & Nilsson, 2021; Matsuoka, 2022; Ng, 2007). The margins are low in hospitality – every bit of additional increment of revenue extracted from the customer has a great impact on the bottom line (Dhurkari, 2023; Leoni & Nilsson, 2021). Revenue management (RM) was originally utilized by airlines, to optimize free inventory, and is nowadays widely used by other industries too. A prime example of this is the hotel industry (Alrawadieh, Alrawadieh & Cetin, 2021; Matsuoka, 2022; Petricek, Chalupa & Levickova, 2022). Having a fixed inventory of a perishable product, possibility of advance sales, uncertain demand, and heterogeneous, segmentable customers, who are willing to pay different prices at different time points practically mandates the usage of RM, which supports in optimally selling said inventory inside a fixed time period (Ivanov, 2014; Ivanov, Del Chiappa & Heyes, 2021; Klein et al., 2019). The majority of hospitality organizations utilize RM in order maximize net revenue and gross operating profit, by selling the right number of the right rooms, to the right customer mix with the right price at the right time, and through the right distribution channel mix (Alderighi et al., 2022; Kool, Westerlaken & Suleri, 2022; Vives & Jacob, 2023). Bayoumi et al. (2013, p.2) define RM as "the science of managing a limited amount of supply to maximize revenue, by dynamically controlling the price/quantity offered".

Pricing and capacity decisions are produced in the background by increasingly complex algorithms (Bodea & Ferguson, 2014; Zhang & Weatherford, 2016). These algorithms adjust prices, by reacting to varying levels of supply and demand (Wamsler et al., 2022). Modern technology enabled RM systems are essential decision support tools for all hotels as they allow effective and efficient demand forecasting, inventory control and pricing policies (Binesh, Belarmino & Raab, 2021; Millauer & Vellekoop, 2019; Zaki, 2022). Recently RM systems have begun to employ customer reviews as an input (Abrate & Viglia, 2016; Viglia & Abrate, 2019; Illescas-Manzano et al., 2023). The RM systems provide recommendations to managers, who after providing their own expert input, distribute the capacity and pricing policies across the multiple distribution channels the hotel uses (Ivanov, 2014; Ng, Rouse & Harrison, 2017; Vives, Jacob & Aguilo, 2019).

As inputs the RM systems use internal and external factors. Internal factors include available capacity, customer segments, current demand, booking curve, historical sales data and the desired distribution channel mix (Ivanov, 2014; Vives & Jacob, 2023; Vives et al., 2019). They also include the price determinants, such as information about the site (e.g., location), hotel type (e.g., size, age), amenities (e.g., breakfast) and property characteristics (e.g., pool) (Alderighi et al., 2022; Bigne, Nicolau & William, 2021; Vives & Jacob, 2023). External factors include season (on a week, month and year level) and possible events (Vives, Jacob & Payeras, 2018). Furthermore, they also include information about the competition and economic context, and quality signal data, e.g., star rating, online rating, chain and brand affiliation (Alderighi et al., 2022; Bigne et al., 2021; Yang & Leung, 2018). And finally, they include reputational information, e.g., online comments (Cantallops & Salvi, 2014; Alrawadieh et al., 2021). As outputs from the RM system, for example, support for the following tactical decisions is received: dynamic prices, capacity controls, offers, overbooking policies, distribution channel mix and length-of-stay limitations (Ivanov, 2014; Ivanov et al., 2021; Vives & Jacob, 2023). RM systems price discriminate over time, while also applying capacity controls (Ivanov & Zhechev, 2012). One of the most common and fundamental RM outputs are dynamic prices (Talon-Ballestero, Nieto-Garcia & Gonzalez-Serrano, 2022).

2.2 Dynamic pricing

> "If effective product development, promotion and distribution sow the seeds of business success, effective pricing is the harvest. Although effective pricing can never compensate for poor execution of the first three elements, ineffective pricing can surely prevent those efforts form resulting in financial success" (Nagle & Holden, 1995, p.1)

Price is the route revenue comes into the company, and pricing decisions should always support the company's objectives (Sánchez-Lozano, Pereira & Chávez-Miranda, 2021). Price is the only marketing mix element that produces revenue, but has received the least attention in research (Dhurkari, 2023; Han & Bai, 2022). The role of pricing is to provide an accurate measure to the customers' product valuation and thus maximize profits (Abrate & Viglia, 2016; Leoni & Nilsson, 2021).

Pricing methods can be divided into the following categories: value-, cost-and competition-based (Dhurkari, 2023; Kool et al., 2022). Of these, cost- and competitor-based are prevalent (Boccali et al. 2022). Cost-based is implemented by evaluating the production cost and adding a mark-up on that, competitor-based by using a competitor's price as a reference. Value-based means evaluating customer value perception and embedding that into the price. Of these the value-based methods has the most potential to improve a company's performance, due to the cost and complexity associated with collecting deep information about the customers' perceptions (Boccali et al. 2022).

Price is the most important output of the RM systems (Petricek et al., 2020). The right price is the one the customer is willing to pay at that timepoint (Ivanov, 2014). The use of dynamic pricing (DP) leads to higher revenue, through the charging of price premiums and added competitive advantage (Binesh et al., 2021; Gao & Bi, 2021; Guizzardi, Mariani & Stacchini, 2022). It enables companies to reduce sold units while increasing revenue at the same time (Han & Bai, 2022; Kool et al., 2022). Using DP prevents setting too high prices when the demand is sluggish, and too low prices during periods of high demand (Vives et al., 2019; Wamsler et al., 2022). If done correctly, DP enables deals for patient customers while enabling companies to attain higher revenues in the meanwhile, much higher than could be attained with fixed price policies (Abrate & Viglia, 2016; Talon-Ballestero et al., 2022). It mitigates two types of risk for the company: the risk of losing revenue, or the risk of having unused inventory, spill or spoilage (Ng, 2007; Cleophas, Kadatz & Vock, 2017).

The decline of brick-and-mortar travel agencies, the rise of online ones, the increased impact of user-generated content and competition, cheaper and better technology and dedicated software have led to DP to become a commonplace practice in hospitality (Abrate, Sainaghi & Mauri, 2022; Neubert, 2022; Ng et al., 2017). Higher classed luxury properties are most successful and most active when it comes to DP because of higher potential gains and larger resources (Alderighi et al., 2022; Binesh et al., 2021), but Prakash & Spann (2022) suggested that for luxury products price variations should be minimized to protect the brand. Abrate, Fraquelli & Viglia (2012) found that 90% of the prices in their observation period were changed in their study of a thousand hotels in eight capital cities in Europe over a three-month period. Mohammed, Guillet & Law (2019) studied last-minute bookings and prices, and every one of their subject properties had adjusted their price at least once in their 7-day observation window. On the other hand, Cirer-Costa (2022) discovered that holiday hotels do not employ DP – as the transient segment is not important enough in that scenario.

2.2.1 *Dynamic pricing principles*

DP is the practice of changing prices over time based on changes in supply (in the form of remaining inventory for the period) and demand, in order to maximize the achieved net revenue (Ng, 2007; Ibrahim & Atiya, 2016; Alderighi et al., 2022). It has three different forms it can take: peak-load pricing, second degree price discrimination and intertemporal price discrimination (Alderighi et al., 2022; Guizzardi et al., 2020). Peak-load pricing means controlling the demand by raising prices when the demand is high. Second degree price discrimination means pricing by segment, for example giving discounts to certain groups at specific times. Intertemporal price discrimination means charging different prices at different timepoints of the same product from the same customers (Alderighi et al., 2022; Guizzardi, Ballestra & D'Innocenzo, 2022).

Different prices can be set for different segments or at different time points, utilizing demand and elasticity information as an input to dynamic prices (Guizzardi et al., 2020; Vives et al., 2019). Abrate, Nicolau & Viglia (2019) separate price discrimination into intertemporal and behaviour-based – charging a different price based on past or potential purchase history. The same distinction is done in Guizzardi et al. (2022), Kool et al. (2022) and Neubert (2022). Behaviour-based discrimination can take two opposite forms: it can either reward existing customers for loyalty, or it can focus on customer acquisition and reward new customers for choosing the company (Kool et al., 2022). This paper adapts and adopts the view of Alderighi et al. (2022), Guizzardi et al. (2020), Zaki (2022) and Abrate et al. (2019) of DP being based on time, demand and behaviour - focusing on intertemporal price discrimination.

2.2.2 *Dynamic pricing and time*

Customers often pay different prices depending on when they book (Ivanov, 2014; Leoni & Nilsson, 2021; Petricek et al., 2020). Being able to book in advance when they want provides customers with one of two benefits: economic (lower price) or security (being sure of getting a room). On the other hand, it exposes them to the risk of either paying too much if price might go down later, or not getting a room at all if the purchase is done too late (Bigne et al., 2021; Guizzardi et al., 2020).

Reacting to the temporally changing demand is what separates DP from purely cost-based pricing (Melis & Piga, 2016). Hotels maximize revenues by adjusting prices as the date of stay draws closer (Chen & Schwartz, 2008; Vives et al., 2018). Better demand forecasting leads to better pricing decisions, and companies can charge significant price premiums by exploiting different booking times and estimating their own market power in real time (Abrate & Viglia, 2016; Vives et al., 2019).

Intertemporal price variations can manifest by prices going up or down, and are based on the sales history, fluctuating demand and the time distance between booking and consumption (Abrate & Viglia, 2016; Guizzardi et al., 2020; Melis & Piga, 2016). The shape of the booking curve and number of available rooms are also factors (Bigne et al., 2021; Ivanov, 2014). The direction and size of these price variations depend on available inventory and the customers' willingness to pay (Leoni & Nilsson, 2021; Vives & Jacob, 2023). Intertemporal price reductions are often of two types: discounts that occur when the date of stay is far, or discounts when it very close (Kool et al., 2022; Vives & Jacob, 2023). Price variations are driven by the type of customer expected to stay at the hotel on each weekday, and the price curve is a parabola, an inverted U-shaped curve, over the booking window (Chen & Schwartz, 2008; Dilme & Li, 2012; Noone, 2016; Klein et al., 2019).

The balancing act between maximizing revenue by increasing prices and risking unsold products leads to hotels adjusting prices constantly, and the likelihood of a price change increases the closer the date of stay is (Petricek, Chalupa & Melas, 2021; Vives & Jacob, 2023). Bigne et al. (2021) claim that the median booking time point is 28 days before the date of stay. Generally, from early on in the booking window prices are likely to go up, and during the last three weeks discounts start to dominate, to stimulate demand in order for hotels to fill up their rooms (Melis & Piga, 2016; Zaki, 2022). If – however – the dates of stay are inflexible for many customers, due to for example special events or high season, the prices can be expected to rise during the days before the date of stay (Chen & Schwartz, 2008; Melis & Piga, 2016; Vives et al., 2019; Petricek et al., 2020).

Leoni & Nilsson (2021) found that in the sharing economy context (Airbnb), last minute price increases actually reduce revenue. Abrate et al. (2012) and Roma, Panniello, & Nigro (2019) found that during weekdays – when the less price sensitive business travellers stay - the lowest prices occur right before the stay, whereas on the weekends, that are dominated by leisure travellers, prices increase right before the date of stay. On the other hand, Mohammed et al. (2019) noticed that properties change prices most frequently on Saturdays, trying to discount the unsold rooms for the more price sensitive leisure market. Earlier research done on the US market (Lee, 2011; Lee et al., 2011) provides some contradicting results by claiming that all weekdays have the same demand.

A rising intertemporal price is driven up by scarcity, perceived risk of missing out and depends on if the customer is willing to wait or not, the customers' degree of patience (Abrate et al., 2019; Chen & Schwartz, 2008; Melis & Piga, 2016). Customers can be split into two categories when it comes to their purchasing behaviour – myopic customers who purchase when the price is lower than they expect, and strategic customers who purchase when they think the price is at its lowest (Binesh et al., 2021; Chen & Chen, 2015). Strategic customers using mobile booking at the last minute may use the DP principles against the company to get a better rate (Neubert, 2022; Yeoman, 2016).

The classic hotel segments – business and leisure - have a different willingness to pay, their elasticity, reactiveness to price change, is different (Kool et al., 2022; Petricek et al., 2020; Vives et al., 2019). The same can be said for the other classical hospitality segmentation – transient versus group customers (Petricek et al., 2020; Talluri, Van Ryzin & Van Ryzin, 2004). In general, the customers can be segmented a third way that covers both of the above: elastic and inelastic (Vives & Jacob, 2023; Vives et al., 2019). Practitioners are able to estimate elasticities with benchmarking, using competitor prices and customer reviews as proxies (Bowie & Buttle, 2004; Alderighi et al., 2022; Talon-Ballestero et al., 2022). Guizzardi et al. (2022) researched in their study the impact of quality perceptions on DP, from a behavioural perspective, and found that the influence of quality on price increases with time. Quality is more important when the booking date is far away, as elasticity changes from inelastic to elastic over time. Estimating the elasticity of each customer allows for more precise segmentation, improving the RM performance (Vives et al., 2019; Vives & Jacob, 2023). While the individual customer's price elasticity is of paramount importance to hotels, they also have to pay attention to cross-elasticities between different customer segments, as well as their cross-elasticities with their competitors (Vives et al., 2019; Xu et al., 2019; Zaki, 2022). How the hotel's prices change will impact competitors' sales is important, as is how different segments react to prices given to other segments.

2.2.3 Dynamic pricing and competition

The hotel industry is considered an intensely competitive market (Rita et al., 2022; Vives et al., 2018). A company can face three types of competition: from alternatives (other companies), from itself (pushing the sale for later) and from within the customer (they can choose to not buy at all). Risk/loss aversion within the customer is more powerful than their feeling of gain from the purchase. Pricing should attempt to mitigate this internal risk whenever possible. If any of these three competitors win – the company can face pressure to reduce the price (Ng, 2007). In the end, the upper price limit is constrained by competition and the lower by cost (Barrows, Powers & Reynolds, 2012).

The more in demand a hotel is the more likely it is to engage in DP (Melis & Piga, 2016; Mohammed et al., 2019). Hotels can charge significant price premiums in real time, based on the demand at booking time and the competitive situation (Abrate & Viglia, 2016; Mohammed et al., 2019). When a direct competitor's inventory decreases, a hotel can raise its price (Sánchez-Pérez, Illescas-Manzano & Martínez-Puertas, 2019). Also, as competition is more intense over the weekend, the prices for hotels go up, especially among hotels with a similar classification (Abrate & Viglia, 2016; Abrate et al., 2012).

DP has to reflect the current demand, competition, season and occupancy – so room prices are changed over the booking window – the time between current date and the date of stay (Alderighi et al., 2022; Chattopadhyay & Mitra, 2019). Relative ease of comparing prices has increased the impact of competitors prices on demand (Gerpott & Berends, 2022). In practice hotel prices are often market driven – hospitality organizations react promptly to changes their competitors make. This can lead to a price trap, a race to the bottom (Avlonitis & Indounas, 2005; Demirciftci & Belarmino, 2022; Ivanov, 2014). To break this the hotels must differentiate through branding and reputation as best they can, or risk turning into a commodity (Carroll, 2011; Kimes, 2011; Vives et al., 2018). Hotel category can act as a mediating factor though, hotels with a high category can maintain a stabler price policy when demand is low, while making sharper price increases when demand is high. High category also allows hotels more independence, as following competitor prices is not as important as it is for budget hotels (Abrate et al., 2012; Sánchez-Pérez et al., 2019).

2.3 Online travel agencies and dynamic pricing

Hotels have a multichannel distribution strategy of using online and offline channels to sell their products to different segments, to maximize revenue and market share while minimizing cost (Raad, Sharma & Nicolau, 2023; Yang & Leung, 2018). Hotels can sell through direct sales, using the global distribution system (GDS), online travel agencies, traditional travel agencies and tour operators (Janssen, 2021; Raad et al., 2023). Due to the perishability of hospitality products: effective distribution is a key factor in their sales (O'Connor, 2020; Petricek et al., 2021; Raad et al., 2023).

The majority of hotels nowadays cooperate with online travel agencies (OTA) to distribute their rooms, who act as intermediaries between hotels and customers (Gao & Bi, 2021; Ivanov et al., 2021; Raad et al., 2023). Expedia.com from the US was the first OTA, founded in 1995 and they now, together with Booking.com, control 75% of the online travel sales market (Raad et al., 2023). A disruptive actor for hospitality, OTAs are the most used distribution channel for transient customers within the European Union, nearly 70% of rooms sold to that segment are through them and now play a critical role for hotels, impacting room prices and profit margins (Gao & Bi, 2021; Han & Bai, 2022; Kim et al., 2023; O'Connor, 2020). The OTA market was worth $432 billion in 2020 (Kim et al., 2023), and control nearly half of the global online travel market (O'Connor, 2020). The emergence of OTAs has multiplied the number of customers who experiment and evaluate dynamic pricing. Pricing is the lever used to better match supply and demand in two-sided markets (Leoni & Nilsson, 2021).

2.3.1 Cost of cooperating with OTAs

OTAs add value while serving as an intermediary on a two-sided market where hotels gain new customers and customers gain from lowered search costs (Janssen, 2021; Raad et al., 2023). The relationship between a hotel and an OTA can show signs of both cooperation and competition (Raad et al., 2023). While hotels and OTAs cooperate to dynamically set the pricing policies together, and hotels feed their pricing policies to the OTAs (Melis & Piga, 2016; Raad et al., 2023; Shadiqurrachman, Ridwan & Kusuma, 2019), the cooperation does create an environment where competition is intensified as there are more companies to compete with and better-informed customers able to choose from more alternatives with little effort (Abrate et al., 2012; Melis & Piga, 2016; Soler et al., 2019).

OTAs are vital to hotel sales, but the emergence of OTAs has increased hotel companies' concerns about margins, the strategic aspect of pricing and increased channel management costs (Erdem & Jiang, 2016). Due to high commissions charged for every sale, hotels would prefer to sell through their own sites, but OTAs have economies of scale on their side – access to multiple hotels for the customers, and access to a plethora of customers for the hotels (Kimes, 2011; Barrows et al., 2012; Yang & Leung, 2018). The commissions charged can be between 15-30% of the revenue sold through the OTA (Ivanov et al., 2021; Neirotti, Raguseo & Paolucci, 2016; Raad et al., 2023). Additional commission payments to an OTA improve visibility on their websites (Ling et al., 2015; Ling, Guo & Yang, 2014).

As the use of OTAs leads to increased sales volume and visibility, so the commission can be understood as a marketing cost (Ivanov, 2014; Nicolau & Sharma, 2019; Raad et al., 2023). OTAs provide a marketing service and charge a price for it (Ling et al., 2014). For larger properties the OTAs clearly bring more volume, while smaller properties also gain in lower marketing costs per customer caused by the increased visibility on the OTA sites (Ivanov, 2014; Inversini & Masiero, 2013; Soler et al., 2019). For smaller properties it can make more financial sense to cooperate with an OTA, than would to be chain affiliated (Ivanov et al., 2021; Raad et al., 2023). OTAs also serve to reduce administrative costs, lower the barrier to enter the online market and provide translation and payment services (Raad et al., 2023).

2.3.2 *The impact of OTAs*

The increased usage of online channels has boosted the usage of dynamic pricing, as updating prices in real time, managed through automation, has become accessible, reliable and cost efficient (Abrate et al., 2022; Bayoumi et al., 2013; Cirer-Costa, 2022; Ibrahim & Atiya, 2016). Online markets have higher competitive dynamics (Gerpott & Berends, 2022; Sánchez-Lozano et al., 2021) which do not accommodate static pricing, online channels act as a natural host to dynamic pricing (Gerpott & Berends, 2022). Increased transparency, competition, number of customers, greater uncertainty in demand, combined with dynamic pricing has created a sort of a self-amplifying feedback loop. The factors mentioned require dynamic pricing, which in itself amplifies the factors mentioned.

The emergence of OTAs and third-party travel forums has led to increased transparency in the market. Customers are able to see online different prices for the same hotel, which depend on for example reviews, booking conditions, date of stay, room type, amenities offered, customer type and how much in advance the customer is booking (Alderighi et al., 2022; Gavilan, Avello & Martinez-Navarro, 2018; Heda, Mewborn & Caine, 2017). The increased transparency has shifted power away from the hotels towards the customer. OTAs provide customers with their own promotions, more alternatives and faster transactions (Raad et al., 2023). As customer comparisons are easier, cross-channel price parity has become an issue for all parties (Guizzardi et al., 2020; Inversini & Masiero, 2013; Janssen, 2021). As it advances the perception of fairness and brand value maintenance, as well as reduces market volatility, the goal for all parties is to show the customers the same price regardless of the distribution channel (Inversini & Masiero, 2013; Nicolau & Sharma, 2019; Talon-Ballestero et al., 2022). On the other hand, some European countries (Germany, Italy for example) have banned rate parity agreements (Nicolau & Sharma, 2019). While they help avoid price wars (Sánchez-Pérez et al., 2019), rate parity agreements have been shown to cause higher market prices overall (Nicolau & Sharma, 2019), and decrease the autonomy of hotels' pricing decisions (Raad et al., 2023).

Transparency can be a constraint to organizations' pricing capabilities, as OTAs attract more savvy, strategic and price sensitive customers (Bigne et al., 2021; Ropero, 2011) The emergence of OTAs has increased the customers' price elasticity (Ivanov, 2014; Vives et al., 2018; Soler et al., 2019). The increased transparency can also lead to – while increasing revenues in the short run – decreased revenues in the long run, due to customers perceiving DP as unfair, resulting in lower satisfaction and loyalty as well as brand dilution (Alderighi et al., 2022; Inversini & Masiero, 2013; Zaki, 2022). Priester, Robbert & Roth (2020) studied, from a behavioural perspective, how customers see individual pricing and found that personalized pricing is seen as unfair, more so when based on location versus based on purchase history. DP and price discrimination policies are tolerated by the public if they are concealed at least partially or are not otherwise seen as unfair, as customers nowadays are sophisticated enough to understand the impact of demand on prices (Kim et al., 2023; Neubert, 2022; Wamsler et al., 2022).

2.4 Online reviews – the rise of eWOM

> “eWOM is any positive or negative statement made by potential, actual, or former customers about a product or company, which is made available to a multitude of people and institutions via the Internet” (Hennig-Thurau et al., 2004, p. 39)

User-generated online reviews – a form of electronic word of mouth (eWOM) - are increasingly used by customers as information sources, to reduce perceived risk and uncertainty before making a purchase decision, facilitating the decision-making (Abrate, Quinton & Pera, 2021; Abdullah et al., 2022; Liu et al., 2022; Wang, Kim & Kim, 2021). After search engines, online reviews are the second most used information source online (Yang, Park & Hu, 2018).

The same factors that have driven the adoption of revenue management and dynamic pricing, and made OTAs the biggest booking channel for the transient segment are increasing the importance of eWOM: larger number of companies and customers, increased transparency and competition in the market, digitalization through cheaper and more developed technology, combined with the ever-fickle demand have made information abundant and at the same time a vital resource for both companies and customers alike (Abdullah et al., 2022; Kim & Hwang, 2022; Mukhopadhyay et al., 2023; Roy et al. 2022).

Online reviews have transformed the way hospitality customers evaluate, choose and share critical incidents they have experienced (Abdullah et al., 2022; Hu, Trivedi & Teichert, 2022; Labsomboonsiri et al., 2022). Phillips et al. (2017, p.1) go as far calling online reviews “the most important information source for customer’s decision making” for hotel purchases. Phillips et al. (2015) state that online reviews coupled with the hotel’s attributes should be thought of as hotel performance determinants. Bi et al. (2019, p.460) call it a “wisdom of crowds” phenomena, while Huang et al. (2022) call it a key asset to a hospitality company, right alongside brand equity. EWOM can come in many forms like blogs, social media posts, photos, reviews (Williams, Ferdinand & Bustard, 2020), but from now on the paper uses the words online review and eWOM interchangeably.

The intangibility of services makes the decision-making process more complex, and reviews are especially important in these contexts (Lopes et al., 2023; Mukhopadhyay et al., 2023; Ye et al., 2023). In the hospitality industry the inseparability of production and consumption causes difficulty to evaluate them in advance, which further increases the importance of eWOM (Chen et al., 2021; Lopes et al., 2023; Mukhopadhyay et al., 2023; Nieto-Garcia et al., 2019).

For the company-customer dyad there has been a clear shift in power towards the customer due to the growing impact of OTAs, social media and eWOM (Neirotti et al., 2016; Huang et al., 2022; Iranmanesh et al., 2022). For hotels the strategic traditional star rating, advertising and official travel office information are losing significance to a more fluid, constantly updated and tactical online reviews which customers find more credible than traditional marketing communications (Castro-Lopez et al., 2022; Huang et al., 2022; Majumder, Gupta & Paul, 2022). In dynamic pricing contexts review rating already trumps hotel classifications as a price determinant (Arora & Mathur, 2020: Sánchez-Pérez et al., 2019). Effectiveness of traditional marketing channels is diminishing (Todri, Adamopoulos & Andrews, 2022; Wilk, Lambert & Meek, 2022). Gursoy (2019) and Wang et al. (2021) note though, that other information sources are used in parallel to eWOM, and earlier, Noone (2016) recommends for companies leverage together both eWOM and company generated content, combining reviews with professional pictures and copy-written descriptions of hotels. It is important to highlight that as eWOM is readily available, for a long time, advertising should be honest and in line with the public opinion expressed online (Zhang et al., 2023).

In hospitality the emergence of a plethora of rich, detailed and personal user-generated content has transformed the way customers seek information about products (Abdullah et al., 2022; Lopes et al., 2023). Customers are now accustomed to having immediate access and up-to-date, accurate information to facilitate information seeking and comparison (Gursoy 2019; Huiyue, Peihan & Haiwen, 2022). Already Mauri & Minazzi (2013) found that 75% of customers read reviews before making a purchase decision on a hotel, while Jiang et al. (2021) claim that figure to be 90%. In online retail Majumder et al. (2022) state that 93% of customers read reviews before purchasing, while in hospitality Akhtar et al. (2019) report that 90% of customers use eWOM before making a purchase decision.

Reading reviews online reduces uncertainty and the cost associated with the search – while providing knowledge to the customer (Majumder et al., 2022; Neirotti et al., 2016). OTAs and third-party travel sites like Tripadvisor provide customers with ample, fresh information hotels and destinations to support in their decision-making, and Tripadvisor dominates over OTAs when it comes to review credibility (Gursoy 2019, Kim & Hwang, 2022; Nieto-Garcia et al., 2019). The higher trustworthiness of Tripadvisor reviews stems from it being a dedicated community-based review site, whereas OTAs are based on transactions, so their reviews are seen less credible (Xiang et al., 2017). On Tripadvisor customers share how their expectations were met, and give an overall rating for a hotel, as well as rate separately the location, cleanliness, service and value for money they have experienced (Castro-Lopez et al., 2022; Neirotti et al., 2016;). Users are able to filter and rank reviews as well on sites where reviews are posted (Xiang et al., 2017).

As stated by the expectation-disconfirmation theory: customers are satisfied if the experienced quality meets or exceeds their expectations (Xu, 2019). Both price and reviews serve as expectation antecedents. If a price or praise is high, then are the expectations. If promises in the form of stellar reviews or high star ratings over-promise, the expectations can be difficult to meet, and failing to meet them has a higher impact on satisfaction than meeting them (Geetha, Singha & Sinha, 2017; Rita et al., 2022). Exceeding the expectations leads to a positive disconfirmation, and failing to meet them leads to a stronger negative one, both evaluations comparing the actual state to the desired state, which will illicit emotions. Meeting or failing to meet expectations causes emotions in the customers, that affects their satisfaction (Zhu, Lin & Cheng, 2020). The reviews following consumption reflect these emotions, not only of the quality perceived, but also of the value perceived, the difference between the price paid and quality (Abrate et al., 2021; Ye et al., 2023; Zhu et al., 2020). Extreme emotions are more likely to illicit a review (Mauri & Minazzi, 2013; Zhao et al., 2015). The review captures the sentiment, concerns and praises of the customer, and these reviews can then be read, and the hotels evaluated based on them, by other potential customers all over the world (Castro-Lopez et al., 2022; Roy, 2023; Tang & Kim, 2022). Reviews act as recommendations to future customers (Majumder et al., 2022; Zhu et al., 2020).

Hu et al. (2019) distinguish between three different avenues of hospitality eWOM research: review helpfulness, impact on customers and impact on companies, Cantallops & Salvi (2014) literature review split the research on the antecedents and impacts of eWOM, on company and customer level. In this paper, however, the split used by Mukhopadhyay et al. (2023) is adopted, into eWOM and performance, eWOM attributes and eWOM and customers.

2.4.1 Characteristics of a review

> "Customer reviews reflect the way consumers describe, relive, reconstruct, and share their experiences." (Xiang et al., 2015, p. 122)

Litvin, Goldsmith & Pan (2008) use a two-dimensional typology for eWOM which is still valid. The dimensions they use are communication scope and interactivity. Communication scope means– one-to-one (e.g., emails), one-to-many (e.g., review sites) or many-to-many (e.g., virtual communities). Interactivity means asynchronous, one-way, communication (e.g., review sites) or synchronous, two-way, communication (e.g., chatrooms). This paper focuses on one-to-many asynchronous eWOM.

Review content can be crudely classified by the information format: unstructured and structured. The structured format includes numerical ratings, an overall evaluation of the experience on a fixed scale, for example from one to five. On some sites ratings can be given to individual aspects of the experience, like cleanliness on Tripadvisor (Tran, 2020). The unstructured textual content of reviews reveals more detail of the customer experience, their emotions and attitudes, the causes behind the structured ratings (Xu et al., 2017; Zhu et al., 2020). The unstructured content can sometimes also include images, either from reviewers or contextual images created by the hotel. These images enhance recall and differentiate products (Tran, 2020).

The three main structured features - rating valence, variation and review volume - have received a lot of attention from researchers (Abdullah et al., 2022; Majumder et al., 2022; Zhang et al., 2019). It is noteworthy that valence in articles is often proxied with the overall rating of the service (e.g., Abdullah et al., 2022; Nieto-Garcia et al., 2019; Yang et al., 2018; Zhang et al., 2023), but it refers to the sentiment polarity of the textual comments (see e.g., Phillips et al., 2017). Ratings are the average evaluation of the rated feature (e.g., "location" on Tripadvisor), so they reflect a large number of opinions in dense format and often function as a straightforward proxy for satisfaction in empirical work (Boccali et al. 2022; Nie et al., 2020; Zhang et al., 2023).

Mauri & Minazzi (2013) provide a framework for analysing reviews, that consists of six factors:

- Valence – are the reviews positive or negative?
- Volume – how many reviews are there?
- Velocity – how many reviews are there in a period?

- History – for how long have the reviews been posted?
- Role – how important the reviews are for the decision-making process?
- Credibility – how trustworthy are the reviews?

Zhao et al. (2015) devised their conceptual framework with review valence, volume, helpfulness, extent, timeliness, and added the source attribute of reviewer expertise – and found that all of these factors significantly positively impacted customers' booking intention. For Zhao et al. (2015) valence and volume are the same as above, helpfulness is the degree customers believe the review assists them in the decision-making process. Extent in Zhao et al. (2015) was defined as how detailed the review is, timeliness as how up-to-date the review is and expertise as how convinced the customers are about the person who wrote the review. In similar vein both Gursoy (2019) and Zeng et al. (2020) provide the ingredients of a review as valence, volume, rating, source and helpfulness. Abdullah et al. (2022) use valence, volume and variation (how different the ratings are among one another).

Review helpfulness is the extent customers think their decision-making is supported by the review. In the hospitality context helpfulness has three components: hotel-, reviewer- and review-related features (Li & Zhang, 2022; Qiao et al., 2022; Zhang et al., 2023). Hotel-related means information about the location, rooms and service, as well as star classification or online rating. Reviewer-related means how well the reader can identify with the person who wrote the review, and how their expertise, engagement and trustworthiness were perceived. Review-related means the length, information richness, readability and valence (Li & Zhang, 2022; Qiao et al., 2022; Roy et al. 2022). Majumder et al. (2022) continue with stating that review rating, length, subjectivity and readability combined with source credibility ae factors determining the review helpfulness. High quality in review texts means they are understandable, informative, relevant and sufficient, providing better support for decision making under uncertain circumstances (Zeng et al, 2020), and quick to grasp categorical measures to facilitate understanding (Zhao et al., 2015). Review readability – how easy it is to understand the text – has a direct impact on sales as it makes the review more helpful, credible and persuasive (Filieri et al., 2019; Majumder et al., 2022).

2.4.2 Impact on companies

Online reviews are widely used by both customers and companies in hospitality (Huiyue et al., 2022; Sann et al., 2021). Online reviews' benefit for the hospitality industry is undeniable, as reviews are increasingly related to a company's performance (Huang et al., 2022; Lopes et al., 2023; Mukhopadhyay et al., 2023). They have been shown to impact sales, occupancy, performance and prices (Abrate et al., 2021; Tang & Kim, 2022; Zeng et al, 2020; Roy et al. 2022; Ye et al., 2023).

In their meta-analysis Hu et al. (2019) found that the median elasticity between ratings and average hotel prices was evaluated to be 0,851 – meaning that an increase of 0,85% in the aggregate review rating translates to a 1% price increase on the hotel's side. Yang et al. (2018) similarly found that in literature the average elasticity between review volume and performance is 0,055, while the marketing field in general had a valence (rating) elasticity of 0,41 and volume elasticity of 0,24. This would imply that in hospitality because of the higher valence elasticity, reviews matter more than for other industries. It also implies the conclusions of Raguseo & Vitari (2017) and Wang et al. (2021) – that brands still affect in hospitality. Raguseo & Vitari (2017) found that branded and non-branded hotels see different effects from eWOM, showing that branded hotels are shielded from eWOM impact, in both good and bad. Branded hotels did not receive the benefits from increased review volume or valence (Raguseo & Vitari, 2017), or increased eWOM volume (Wang et al., 2021) – thus the lower-than-normal volume elasticity. This is at least partially contradicted by reviews being more relied on, than brand information, on travel apps (Huiyue et al., 2022).

The positive relationship between review rating valence and performance has been studied extensively (see e.g., Alrawadieh et al., 2021; Hu et al., 2019; Yang & Leung, 2018; Yang et al., 2018). A clear, positive relationship between online ratings and sales has been found in many studies (Nieto-Garcia et al., 2019; Phillips et al., 2017; Roy et al. 2022; Sánchez-Pérez et al., 2019). Opposingly, Zhao et al. (2015) found that only negative reviews correlated significantly with a lowered booking intention, whereas positive reviews did not. One of the earliest works found on review valence and hospitality sales, Ye, Law & Gu (2009) found that a rating increase by 10% can improve sales 4,4%, and that a 10% increase in the variance of the ratings can lower sales by 2,8%. Ye et al. (2011) found that rating valence had significant impact on sales, but rating variance did not. In a restaurant setting Wilk et al. (2022) state that a one-star rating improvement can provide 5-9% added revenue.

Hu et al. (2019) found no impact for individual cleanliness or service ratings on room prices in the literature. Phillips et al. (2017) studied various hotel attributes' effect on occupancy rate through the review sentiment and found that room quality, Wi-Fi and positive previous eWOM were the biggest contributors, they also found a significant negative correlation for value for money rating and occupancy rate. In their study the review sentiment analysis was polarity level, reviews being positive (1), neutral (0) or negative (-1). Similarly, Nieto-Garcia et al. (2019) disaggregated review ratings to find that the ratings for location, staff and facilities are the things that boost revenue per available room, and value for money rating was not significant. While Neirotti et al. (2016) found no evidence for it, Huang et al. (2022) state that high variability in online ratings can cause lost sales.

Emotions expressed in review texts are success indicators (Liu et al., 2022). Companies with a positive online reputation are able to charge a price premium and gain a larger average transaction online (Geetha et al., 2017; Sánchez-Pérez et al., 2019). Building on Phillips et al. (2015), Yang et al. (2018) found that review valence improvement allows for higher revenue per room in luxury hotels, whereas an increasing review volume drives up occupation rate in lower class hotels. Similarly, Ye et al. (2023) report that ratings drive occupancy for the low to medium-priced hotels, whereas a rating 7,5 and over meant that ratings drive prices too. If the value of online reviews and ratings increases, it allows improving price positioning – making a hotel able to have a higher base-line price (Abrate & Viglia, 2016; Phillips et al., 2015; Yang et al., 2018). A better online reputation also supports non-financial goals like branding and competitive positioning (Mariani & Visani, 2019).

By studying review quality, quantity, consistency and recency - Xie, Chen & Wu (2016) demonstrated that time heals wounds, stating that the impact of review quality on hotel occupancy lasts at least a couple quarters, whereas quantity, consistency and recency only have short-term implications. This would imply that customers do not read every single review about a hotel, but only some that are the most recent or relevant, for example only the extreme ones with top or bottom rating. This view is supported in Cantallops & Salvi (2014) literature review.

There is, however, some contrary evidence when examining the bottom line, profit impact of reviews. Studies state that valence, volume and variation do not significantly impact the profit of a hospitality company (Abdullah et al., 2022; Neirotti et al., 2016). Abdullah et al. (2022) speculate in a restaurant setting that the cost of serving more customers reduces the margins to an extent that diminishes profit. Neirotti et al. (2016) provide evidence for a declining unit profit margin, and increasing occupancy rates in hotels which do not benefit hotels as much because of strong seasonality and the limited capacity they have. They state that the increased transparency caused by OTAs and travel infomediaries like Tripadvisor does create value, but this is caught by the online operators. Luxury hotels, and properties in low competition locations are able to capitalize better from the increased visibility. They too did find evidence on revenue growth, and close on "better online visibility is more important than greater visibility" (Neirotti et al., 2016, p.1142).

Online reviews can be used to estimate the current competitive landscape, and valuable information asset to companies (Boccali et al., 2022; Gerpott & Berends, 2022; Nieto-Garcia et al., 2019). Reviews can be mined for relevant factors for improving the company offering, as well as more effective service failure recovery (Iranmanesh et al., 2022; Phillips et al., 2017; Raguseo & Vitari, 2017; Xu et al., 2019). Reviews also provide a good, reliable and cost-efficient starting point for both internal and external analysis, as well as benchmarking for hotels (Geetha et al., 2017; Huiyue et al., 2022; Liu et al., 2022;). Furthermore, Mariani & Visani (2019) state that online rating is a better predictor on hotel performance than traditional customer satisfaction. And finally, Boccali et al. (2022) state that reviews can be used as a proxy for customers' value perceptions, and that way provide a relevant, trustworthy and up-to-date basis for implementing value-based pricing.

2.4.3 Impact on customers

Most customers try to avoid risk, and review ratings are a quick heuristic to reduce the risk of the purchase, both from a financial and social perspective, by reducing information asymmetry, enabling comparison and avoiding making the wrong choice (Lopes et al., 2023; Majumder et al., 2022; Mukhopadhyay et al., 2023). Reviews provide quality reassurance in the form of a testimonial (Abdullah et al., 2022; Huang et al., 2022; Labsomboonsiri et al., 2022). Review valence, its variation and the richness of the review content can impact customers purchase intentions positively or negatively (Phillips et al., 2017; Majumder et al., 2022; Zhang et al., 2019).

The process behind the reviews' effect has a functional significance to the readers – it adds to their knowledge another users' experience (Huang et al., 2022; Majumder et al., 2022). According to the elaboration likelihood model of information processing: customers process review information via two routes, central and peripheral (Huiyue et al., 2022; Majumder et al., 2022; Zhang et al., 2023). Which route is applied depends on the type of information received, for example the review text quality is processed the central route and requires cognitive effort. The quality of the argument is assessed, and an evaluation is formed. A rating goes through the peripheral route, and requires less cognitive effort. Some indirect cues like source credibility are weighed, and the information is internalized as a simple evaluation (Huiyue et al., 2022; Majumder et al., 2022; Zhang et al., 2023). The peripheral information is the context of the review, whereas the central is the content, and both matter when thinking how reviews impact customers, and these can be used together to form an opinion of the review (Zhang et al., 2023). What this means when purchasing hospitality products: as the decision is perceived as complex and risky, the purchase decision is done in two stages: information search and evaluation (Baek Ahn & Choi, 2012; Gursoy 2019). Initially in the information search stage customers use heuristics to reduce cognitive effort, using peripheral cues, like aggregate ratings and source credibility, to narrow down their options into a consideration set. In the evaluation stage they use more complicated information, central cues like text and photos, and the quality and the length of the argument, to make their final choice (Baek Ahn & Choi, 2012; Gursoy 2019; Nieto-Garcia et al., 2019). The decision process resembles a funnel, starting wide with a lot of options, and landing on one.

Goods can be separated into search (every-day)- and experience-goods (Baek, Ahn & Choi, 2012; Fang et al., 2016; Majumder et al., 2022). If it is possible to get information about product quality in advance, it is considered a search-good, whereas experience goods are more subjective and the quality is experienced through one's senses upon consumption (Majumder et al., 2022; Mukhopadhyay et al., 2023). The purchase of an experience good, like a hotel room or a vacation, is based on subjective judgement and the information provided by the hotel may not be enough (Fang et al., 2016; Majumder et al., 2022; Zhang et al., 2023). Normally, the hotels cost more and are both planned and consumed over a longer time than other experience goods like dinner (Chen et al., 2021; Filieri, 2016; Gursoy 2019), so the evaluation stage can be an even more complicated experience, leading to more reviews read (Filieri, 2016; Gursoy 2019; Huiyue et al., 2022). To get the most out of this, customers need authentic reviews with both structured and unstructured content (Li & Zhang, 2022; Majumder et al., 2022; Roy et al. 2022). For the more expensive experience goods reviews are utilized more and at the earlier information search phase of the decision process, and the peripheral cues are used to determine the review helpfulness before committing to reading it (Chen et al., 2021; Fang et al., 2016; Zhang et al., 2023). Roy (2023) used sentiment analysis to assess if customers used subjective or objective evaluation when writing reviews, and found that customers in luxury hotels, in line with the elaboration likelihood model, used subjective evaluation, while customers in budget hotels used objective evaluation.

Textual review content has been found to be more impactful on purchase decisions than ratings, and on sales ratings have an indirect effect, acting through the sentiment of the review – the review sentiment is what drives the sales (Hu, Koh and Reddy, 2014; Liu et al., 2022; Majumder et al., 2022). Liu et al. (2022) analysed restaurant review texts, using sentiment analysis results in a regression model, to find that out of the six main emotions in their model (love, trust, anger, joy, surprise and sadness) –love expressed in a review had the strongest impact on purchase decisions, with anger and surprise having significant impacts too. Huiyue et al. (2022) note that the more vivid, descriptive and consistent with previous opinions a review is, the more credible the information provided is perceived. Photos and images add to the vividness of a review (Huiyue et al., 2022; Roy et al. 2022). It is also noteworthy that different review topics can have a different impact, even if the valence is the same (Qiao et al., 2022). Reviews about service failures were found more helpful than complaints about money, for example. And customers in luxury hotels complained about the service more, whereas customers in budget hotels complained about the facility (Qiao et al., 2022).

Reviews are the main social influence impacting online purchase decisions, amplified by similarities between the reviewer and the reader (Castro-Lopez et al., 2022; Gavilan et al., 2018; Huang et al., 2022). Todri et al. (2022) found that geographical distance correlates negatively with the effectiveness of eWOM.

As review are not paid for, customers find reviews and ratings as non-commercial, authentic, independent and trustworthy (Castro-Lopez et al., 2022; Mukhopadhyay et al., 2023; Tang & Kim, 2022; Wilk et al., 2022). It has been shown that financial compensation for reviewing is no match for social compensation like reviewer badges (Labsomboonsiri et al., 2022; Yen & Tang, 2019). The information contained in a review reduces the customers' search cost and is perceived reliable, thus increasing a hotel's trustworthiness in the process, and that has an impact on sales (Kwon, 2023; Majumder et al., 2022). Furthermore, as reviews are posted spontaneously, they lack biases caused by social desirability (Kwon, 2023). Some concerns have been raised about fake reviews, companies boosting their image online or badmouthing competitors (Wang, Fong & Law, 2022; Xiang et al., 2017). Wang et al. (2022) devised a framework for detecting fake reviews, and point out that faking positive reviews is easier than faking negative ones.

A high number of reviews implies popularity (Roy et al. 2022; Zeng et al, 2020), and the increased visibility increases the likelihood that a hotel ends up in the customer's choice set (Zhao et al., 2015). Review volume also implies that review valence is credible (Abdullah et al., 2022; Castro-Lopez et al., 2022; Tang & Kim, 2022). Gavilan et al. (2018) found that volume provides more credibility to negative reviews.

As more and more reviews get posted, customers can get exposed to information overload, which can also make the information harder to process (Dickinger, 2011; Filieri, Galati & Raguseo, 2021; Zhang et al., 2023). It is often not possible for customers to read every single review, so the most recent ones are the most impactful (Liang et al., 2022; Zhang et al., 2023 Zhang et al., 2019). An earlier survey reported that customers read 6-12 reviews before booking (Hospitalitynet, 2014), while also in an earlier work, Hu et al. (2014) found that the most recent reviews were the most impactful for online purchases. Huiyue et al. (2022) state that customers read 5-6 reviews before making a decision.

Prospect theory (Kahneman & Tversky, 1979) describes that the way people weigh their options and makes choices is based on what they can gain and lose. According to it, negative attributes have more impact on satisfaction than positive ones. This would indicate that negative reviews are perceived as more relevant, and more helpful (Xu, 2019). Also, as customers prefer to project a positive social image, there is a smaller volume of negative reviews, perhaps they are perceived as more valuable (Yen & Tang, 2019). Lacking physical contact with the writers, online review readers form impressions based on the language, as well as detailed and factual content, valence and extremity of the reviews (Filieri, 2016). Negative reviews often contain more detail, than positive ones, and are deemed more helpful and credible than positive ones by the customers, especially ones expressing a moderate level of negativity (Huiyue et al., 2022; Qiao et al., 2022; Zhang et al., 2023). Opposingly, Xiang et al. (2017) found proof that a positive review sentiment is perceived as more helpful. Berezina et al. (2016) found that customers write more about physical, tangible things in negative reviews, while Zhang et al. (2023) point out that as negative reviews contain more details. It might be that positive reviews are helpful for customers looking for heuristic cues to use, and negative reviews for those who are looking for deeper information. They also found that recent positive reviews are found useful, and, when looking at a few reviews one after another, inconsistency is not as negative an impact for negative reviews as it is for positive ones (Zhang et al., 2023).

Contrastingly, reviews with more extreme sentiment are more persuasive (Fang et al., 2016; Majumder et al., 2022), but are experienced less trustworthy due to their apparent emotionality (Filieri, 2016; Zhang et al., 2023). The impact of negative reviews cannot be fixed by even large discounts (Han & Bai, 2022). Mauri & Minazzi (2013) found that reviews exposing both positive and negative details are perceived as more detailed and credible, but Filieri (2016) added that for this to be accurate, a moderate rating should accompany the moderate text. Balanced reviews generate more attention and processing motivation in the reader, making them interesting (Roy et al. 2022). On the other hand, Yang et al. (2018), and Qiao et al. (2022) found that single-sided reviews are perceived as more enjoyable and provide a clearer heuristic, that is more effective in supporting the customer's decision making. All in all – both kinds of messages increase visibility and raise customer awareness (Cantallops & Salvi, 2014; Majumder et al., 2022). In the hospitality context, luxury hotels with a high star rating illicit extreme negative reviews more easily, as they have higher expectations to meet, while authentic service receives most of the extreme positive reviews (Filieri, Galati and Raguseo, 2021). A clear, personalized and timely response from the hotel alleviates the negative consequences of reviews (Lopes et al., 2023; Xu & Zhao, 2022).

2.5 Hedonic price theory

> "The price of a product can be regarded as a function of the measurable utility-affecting attributes or characteristics of the product" (Gibbs et al., 2018, p.47)

Hedonic price theory (HPT) is a theoretical framework originally more used in real-estate management, but now most widely used to obtain determinants of lodging price (see e.g., Abrate & Viglia, 2016; Abrate et al., 2012; Abrate et al., 2019; Andersson, 2010; Dudás et al., 2020; Falk, 2008; Gibbs et al., 2018; Hu et al., 2019; Latinopoulos, 2018; Leoni & Nilsson, 2021; Lozano, Rey-Maquieira & Sastre, 2021; Pawlicz & Napierala, 2017; Rigall-I-Torrent & Fluvia, 2011; Sánchez-Lozano et al., 2021; Sánchez-Pérez et al., 2019; Soler et al., 2019; Vives & Jacob, 2023; Zhang, Ye & Law, 2011).

HPT is based on utility theory, which states that customers purchase products if the perceive the acquired utility is higher than the cost incurred (Lancaster, 1966; Xu, 2019). Originally framed by Lancaster (1966) and developed further by Rosen (1974), HPT assumes that – rather than homogenous entities – differentiated goods and services are bundles of attributes embedded in the product, difficult to repackage, whose contribution is valuated individually. The aggregate of these attributes makes up a linear function for the product's price (Abrate et al., 2019; Sánchez-Lozano et al., 2021; Soler et al., 2019). In Rosen's (1974) model producers adjust their products to include attributes wanted by the customers. For the case of hotel price determinants, the model brings together the market (location, time, season), supply (the hotels) and demand (the customers) who interact through prices (Sánchez-Lozano et al., 2021; Soler et al., 2019). Producers and customers interact in the market and attempt to maximize their profit and utility, respectively (Andersson, 2010; Latinopoulos, 2018; Zhang, Grisolia & Lane, 2023). Hotels seek value by making pricing decisions, customers seek utility by choosing (or not) to purchase.

An assumption of hedonic price models (HPM) is that variable costs are negligible, which, as fixed costs are so high, can be thought to be the case for hotels (Falk, 2008; Soler et al., 2019). Another assumption in HPMs is a transparent, perfectly competitive market, an assumption that also can be thought of as fulfilled for hotels, due to online sale and reviews (Latinopoulos, 2018).

Implied by the theory – the contributions of the attributes on price are estimated through multiple regression analysis, to see the scope and amounts attached to these attributes (Gibbs et al., 2018; Rosen, 1974; Tavares, Tavares & Santos, 2022). Hotel rooms are differentiated by attributes, for example location, hotel size, online reputation and service quality – so HPMs can decompose the individual factors and assign a market price for the different components that make up the hospitality product as a whole (De Olivera Santos, 2016; Han & Bai, 2022; Hu et al., 2019). This market price is a proxy for the customers' expected value and willingness to pay (De Olivera Santos, 2016; Gibbs et al., 2018). Once known, the value of different individual factors can be used for improving the offering, investment decisions and pricing decisions (Arora & Mathur, 2020; De Olivera Santos, 2016; Rigall-I-Torrent & Fluvia, 2011). Attributes valued positively have a positive sign in the resulting regressed equation, and the result parameters can be used to see what attributes hotels should emphasize in their differentiation (Latinopoulos, 2018; Rigall-I-Torrent & Fluvia, 2011; Soler et al., 2019). Supported by Arora & Mathur (2020), Rigall-I-Torrent & Fluvia (2011) highlight that, by extracting attribute prices for existing hotels, entirely new, prospective combinations can be evaluated before any costly hotel projects are initiated.

The hedonic price models are often used to model price setting in empirical contexts, and implemented through observing prices that are collated as panel data (price info over time per hotel) (Lozano et al., 2021; Hu et al., 2019; Soler et al., 2019). The strength of HPMs is that they always use observed, real market data (Espinet-Rius et al., 2018; Gibbs et al., 2018; Soler et al., 2019). The modern dynamic pricing algorithms use the interplay of current price, demand and supply to calculate the new price (Guizzardi et al., 2022). While not the actual purchase price, the observed dynamic, demand-influenced, prices can nonetheless be understood as the real price (Rigall-I-Torrent & Fluvia, 2011; Latinopoulos, 2018; Vives et al., 2019). When prices are observed, the size of the hotel, how many rooms it has, is observed as a proxy for available capacity (Abrate et al., 2012; Guizzardi et al., 2022). The length of the booking window, time from observation date to date-of-stay, is used as a proxy for seasonality (Abrate & Viglia, 2016) which does imply an unrealistic, smooth and linear demand when the regression is implemented (Bigne et al., 2021). Another strength of HPMs is that - as the data is publicly available and the method is simple – it is fast and cheap to implement (Rigall-I-Torrent & Fluvia, 2011). It should be noted that analysing purely the hotels' qualities approaches the problem from the supply side, but prices are also impacted by the demand side by factors like reputation and season on month, week or day level (Arora & Mathur, 2020).

The weakness of HPMs is that the results cannot always be generalized, as the model usually focuses on an area that has internal spatial homogeneity (Andersson, 2010; Soler et al., 2019). The relevant attributes can vary from region to region, and each model is a description of a specific location and its hotels (Arora & Mathur, 2020; De Olivera Santos, 2016; Illescas-Manzano et al., 2023).

Some studies aim for the variables examined to cover as many variables as possible (e.g., Soler et al., 2019) while others choose one attribute to focus on (see e.g., Illescas-Manzano et al., 2023), while controlling a limited number of other variables (Arora & Mathur, 2020). There is a consensus among the HPT literature that hotel category and location factors should be in the models. Otherwise, due to the spatial differences between destinations, there is no consensus on what other variables should be deployed, and the theory does not provide guidelines for variable selection (Gibbs et al., 2018; Latinopoulos, 2018; Soler et al., 2019). To get over this weakness, in this paper, previous literature and the text above will be utilized to provide a choice set of variables, in line with previous literature (Andersson, 2010; Soler et al., 2019). Petricek et al. (2020) and Vives et al. (2019) mention location, hotel class, room type as common examples of variables in HPMs, nonetheless the literature (see e.g., Soler et al., 2019) emphasize that each market is different and in the end the researcher must always choose the appropriate ones for the particular case at hand.

There is also consensus in the literature that, instead of linear regression, a loglinear regression provides better results. This reduces skewness and prevents parameter overestimation in models where there are large values on some variables and, on the other hand, variables with very small absolute values, in the same model (Arora & Mathur, 2020; Espinet-Rius et al., 2018; Illescas-Manzano et al., 2023; Soler et al., 2019).

2.5.1 Hedonic price theory research

Falk (2008) used HPT to investigate what attributes customers appreciate in ski resorts, summarizing the lift ticket prices as a function of a resort's attributes. Andersson (2010) studied hotel price determinants in Singapore, using a survey as a customer feedback variable. Zhang, Ye & Law (2011) were among the first to incorporate eWOM into an HPM, in the form of Tripadvisor ratings as a hotel price determinant. Herrmann & Herrmann (2014) the impact of an event (Oktoberfest in Munich) to the valuation of price components, and found proof of intra-week variation of demand. De Olivera Santos (2016) researched price determinants for hostels worldwide, comparing how the important factors changed with location. Espinet-Rius et al. (2018) studied what components make up the price of a cruise. Gibbs et al. (2018) used HPT to find Airbnb property price determinants in five cities. Latinopoulos (2018) incorporated geographic information system data to an HPM to investigate the impact of a sea view on room price in Greek resort.

Abrate & Viglia (2016) investigated the relative importance of several price impacting factors and their evolution by developing a conceptual model with three types of variables: tangible (physical characteristics), reputational (e.g., ratings and reviews) and contextual (e.g., location, competition, timing, events, demand) which act on two levels, strategic and tactical. Between-hotel price variations reflect strategic choices, while within-hotel price variations reflect tactical choices. Abrate & Viglia (2016) found that star rating impacts more the longer-term strategic pricing, and online rating and amount competitors are more important for the short-term tactical pricing. As hotels have implemented dynamic pricing, they also found that online reviews gain importance and that booking time should be considered an important variable.

Pawlicz & Napierala (2017) investigated prices on three different OTAs and confirmed the impact of star ratings, the OTAs own ratings as well as distance from the centre. They also stated that customers do not know the difference between the official star rating and the OTAs own. Star rating is a variable that explains price variations for both business and leisure travellers. As it is an aggregate of multiple factors - it should not be used together with other service standard characterizations like room size, facilities or complementary services. While the star rating is an important factor in determining both the customers' willingness to pay and their service expectations – it should be noted that organizations have no incentive to acquire a rating higher than their service providing capabilities justify (Pawlicz & Napierala, 2017).

Soler et al. (2019) researched customers' willingness to pay in Algarve holiday hotel market through HPT, and found that the previous day's price, hotel type, star rating and services were the biggest determinants of the prices. Dudás et al. (2020) applied HPT to study how investing in different attributes could be justified, how much a room price can be increased if a pool is built, for example. Sánchez-Lozano et al. (2021) too investigated the market in Spain, namely Madrid, and through collecting data from official sources and Booking.com for multiple variables found that hotels in different classes have different price determinants, and that online rating and hotel size were significant. They recommend using a Tripadvisor rating kind of variable in hedonic studies.

Sánchez-Pérez et al. (2019) used HPT to study how star rating, country location, competition intensity and eWOM rating impact hotel prices in different countries, showing that the importance of these factors is country and star rating dependent, and that positive eWOM has a positive impact on hotel price. A rating increase of one point on Tripadvisor can increase the room rate by 4,6%. Supporting them, Illescas-Manzano et al. (2023) also studied how price determinants for hotels differ between countries in the long term. As in other works, Abrate & Viglia (2016) – they found that the physical core hotel attributes, competition and location are still important, and that while still significant, the impact of star rating is declining due to lack of integration with the customer voice.

Lozano et al. (2021) used HPMs to form a hedonic price function to model price setting and seasonality for hotels in different Spanish regions, noting that the local competition heavily influences the price. Arora & Mathur (2020) examined how price determinants vary between countries in order to create a framework that supports strategic investment decisions – i.e., where to build a hotel? They found that hotels in tourist destinations can charge a significant price premium over similar hotels located in a less of a tourist destination. Furthermore, they found that having a pool, providing free breakfast and higher online rating are positively correlated with price, while size is negatively correlated with average room price.

Table 1: eWOM and price in hedonic studies

Reference	eWOM operationalized as	Price operationalized as
Andersson (2010)	Ratings	Fixed room rate
Zhang, Ye & Law (2011)	Ratings	Annual average price
Rigall-I-Torrent & Fluvia (2011)	None	Fixed room rate
Abrate & Viglia (2016)	Ratings	Dynamic prices from Booking.com over time
De Olivera Santos (2016)	Ratings	Fixed room rate
Pawlicz & Napierala (2017)	Ratings	Fixed room rate
Gibbs et al. (2018)	Volume and ratings	Fixed room rate
Latinopoulos (2018)	None	Fixed room rate
Abrate et al. (2019)	Ratings	Dynamic prices from Booking.com over time
Sánchez-Pérez et al. (2019)	Ratings	Fixed room rate
Soler et al. (2019)	Ratings and Travel Award dummy	Dynamic, but collected on one day, for a 28-day period
Dudás et al. (2020)	None	Fixed room rate
Lozano et al. (2021)	None	Fixed room rate
Sanchez-Lozano et al. (2021)	Ratings	Dynamic, but collected on one day, for one year ahead
Vives & Jacob (2023)	Ratings	Fixed room rate
Illescas-Manzano et al., 2023.	Ratings	Fixed room rate

2.6 Hypothesis formulation

With fixed costs being high compared to variable costs in hospitality, effective pricing is what drives the performance (Vives et al., 2019; Vives & Jacob, 2023), in hospitality effective pricing trickles down to the bottom line (Leoni & Nilsson, 2021; Matsuoka, 2022). While the tangible attributes still matter, online reputation is gaining importance in pricing (Abrate & Viglia, 2016, Vives & Jacob, 2023). The intangible aspects and inseparability of the hospitality product make reviews important to customers (Lopes et al., 2023; Mukhopadhyay et al., 2023; Nieto-Garcia et al., 2019), as they provide a credible recommendation from other customers who have already experienced the service (Castro-Lopez et al., 2022; Huang et al., 2022; Majumder et al., 2022).

Positive reviews have a positive impact on hotels' sales and performance (Illescas-Manzano et al., 2023; Roy, 2023). Good ratings ensure a higher revenue per available room (Yang et al., 2018). Positive reviews have also been found to have a positive impact on the average price hotels charge, so customer understating is vital for the industry (Hu et al., 2019; Petricek et al., 2021; Ye et al., 2011). Positive ratings also impact dynamic prices positively (Abrate & Viglia, 2016; Sanchez-Pérez et al., 2019).

Positive emotion in the textual content of the reviews is a predictor of success (Liu et al., 2022). The ratings provided in the reviews act as a quick heuristic, but it is the emotion expressed in the reviews that closes the sales (Gursoy 2019; Hu et al., 2014; Liu et al., 2022; Majumder et al., 2022). The purchase decision is a subjective one for an experiential product like a hotel room (Majumder et al., 2022; Zhang et al., 2023), and customers require rich information to make up their mind (Majumder et al., 2022; Roy et al. 2022). This causes the emotional intensity of the reviews to be an important factor in pricing too (Nie et al., 2020; Zhang et al., 2023). Reviews are the way this sentiment is shared with future customers (Roy, 2023; Tang & Kim, 2022), the textual content provides the causes behind the numerical ratings (Xu et al., 2017; Zhu et al., 2020).

The preceding 3 paragraphs lead to the first two hypotheses:

H_1: Review sentiment impacts the dynamic price set by a hotel

H_2: Positive review sentiment impacts hotel price positively

The booking time and demand impact the customers' elasticity of demand, and this effect can be observed on dynamic prices (Kool et al., 2022; Petricek et al., 2020; Vives & Jacob, 2023). Factors impacting prices are correlated with time (Guizzardi et al., 2020; Melis & Piga, 2016). Price sensitive leisure customers tend to book earlier or at the very last minute (Ibrahim & Atiya, 2016; Roma et al., 2019), making the price curve resemble an inverted U-shape (Noone, 2016; Klein et al., 2019). The value of tangible hotel attributes go up, and the value quality signals go down when the date-if-stay is closer (Guizzardi et al., 2022; Vives, et al., 2018). As prices go up due to higher demand, so does the fear of missing out in the customers' minds (Abrate, et al., 2019; Melis & Piga, 2016; Ng, 2007). When demand is high hotel prices climb to their maximum height independent of reviews (Herrmann & Herrmann, 2014). As customer perceptions and elasticities shift over time, due to the different segments booking at different time-points, and for different days – the impact of review sentiment on price should be different too. Which leads to two more hypotheses:

H_3: Time moderates the impact of review sentiment on the dynamic price of a hotel

H_4: Demand moderates the impact of review sentiment on the dynamic price of a hotel

Based on the above and the literature review – the initial framework can be amended with the hypotheses, see Figure 3 below.

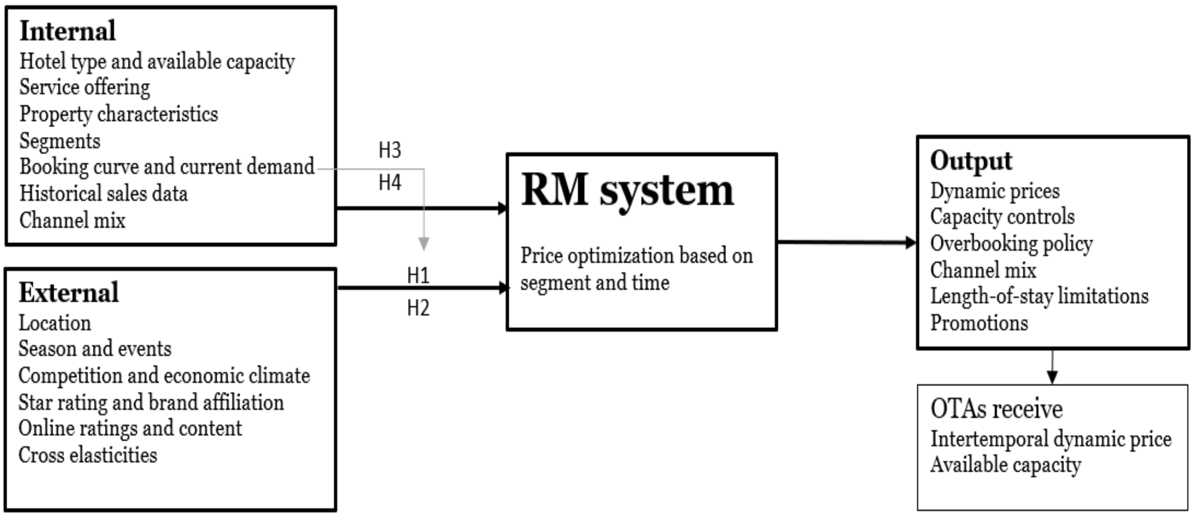

Figure 2: Final framework

3 METHODS

3.1 Research philosophy

The aim of examining pricing and reviews was achieved through content analysis, implemented as an online observational study. The paper approaches the problem from a positivist point of view, the observed variables are perceived to describe the concrete physical reality (Hudson & Ozanne, 1988; Hirschman, 1986). The ontology adopted in this paper dictate that the variables observed and calculated describe what is happening in reality, and that the results derived can be utilized in other contexts (Saunders et al., 2009). The epistemology adopted under the positivistic philosophy of the researcher means that actions of the observed agents and the relationships between them constitute the knowledge observed and produced in this paper. The approach chosen in this study is inductive, as new theory is produced from the observed data. The time-span of the study is longitudinal, as the data used covers a three-month period.

3.2 Design of the study

The aim of the paper was to answer the question: "how does review sentiment impact hotel dynamic prices?". This was done through testing hypotheses H_1, H_2, H_3 and H_4. To do this, hedonic price theory was chosen, as it is a common way to attribute the direction and magnitude of the impact different factors have on price (Abrate et al., 2019; Rosen, 1974). First, data is collected, online in this case, and then used as dependent and independent variables in the hedonic price equation, which can be simplified as:

$$P = f(E, I)$$

In this equation P equals price, E equals a vector of external factors and I equals a vector of internal factors. Multiple multivariate loglinear regressions were then executed in order to see the impact of each attribute (in line with Yen & Tang, 2019). It is noteworthy that this paper studied correlation, because the causation effects cannot be detected with this research design.

3.3 Data collection

The target market chosen was the Helsinki hotel market. Partially due to familiarity, this choice is motivated by the fact that no studies using data the Nordic market in general and Helsinki in particular were found. As each market is considered unique in the hedonic price literature – this paper adds to that body of work (Arora & Mathur, 2020; Illescas-Manzano et al., 2023). Helsinki is the capital of Finland, with an area of 720 km², 217 km² of which is land area. In 2022 the city had 658457 inhabitants, while the Helsinki region had over 1.5 million inhabitants (Tietokeskus, 2022). In 2017 there were 71 hotels in Helsinki, with over 10000 rooms that were utilized 72.9% by over 1.9 million customers. The average, real price acquired was 74.55€ during the observation year (Statistics Finland, 2018).

Data sources for this paper were Tripadvisor.com, Booking.com and Hotels.com (referred to without the URL ending from now on for readability). These three are often sited as sources for multi-platform analyses on eWOM (Mariani et al., 2019). Tripadvisor and Booking are commonly used and recommended in hedonic price models (HPM) (Nieto-Garcia et al., 2019; Pawlicz & Napierala, 2017; Soler et al., 2019). Using real market data is one of the strengths of HPMs, the observed dynamic prices respond to real changes in season, demand and reflect real customer behaviour (Petricek et al., 2020; Soler et al., 2019).

Tripadvisor is the most popular travel website in the world, and it can be assumed to have the largest eWOM impact (Filieri et al., 2019; Qiao et al., 2022; Yang et al., 2018). Sann et al. (2021) site it as the most common source of eWOM in hospitality research. All the information the customer needs for decision making is available there, and it is considered a reliable data source in hospitality (Soler et al., 2019; Yang et al., 2018). Over one billion, adequately long, credible reviews display a broad distribution of sentiment (Li & Zhang, 2022; Tripadvisor, 2023; Yang et al., 2018). Tripadvisor also has strict publishing rules, IP address and email checks and user reporting in place to ensure the quality of the reviews (Liang et al., 2022; Zhao, Xu & Wang, 2019). On top of the textual comments, users can enter a Likert-type grade between one and five for rating various aspects of their stay in the hotel (Hu et al., 2022). Customers first provide an overall rating, and then independently rate location, cleanliness, service and value of the hotel. The independent ratings are not aggregated to make up the overall score (Qiao et al., 2022).

Booking is the largest OTA, and together with Expedia Group control 75% of online travel sold in 2022 (Alderighi et al., 2022; Raad et al., 2023). It boasts having 28 million listings in 2.5 million properties, in 226 countries, and over 1.5 million reservations daily (Booking, 2023; Qiao et al., 2022). Alderighi et al. (2022) and Pawlicz & Napierala (2017) recommend using more than one OTA as a source, as this increases generalizability and the internal validity of the study. Mariani (2020) and Xiang et al. (2017) also recommend using multiple sources of data whenever possible. In this paper Hotels.com was chosen as an additional price source, due to it belonging to the Expedia Group, the other large OTA conglomerate in the industry (Nicolau & Sharma, 2019; Raad et al., 2023).

This thesis was initially started in 2017, and that is when the data was collected. The data was collected between 1.6.2017 – 31.07.2017. Hotel and review data were collected from online sources, and price data was collected by simulating customer booking behavior online. The hotel data was observed from Tripadvisor. The price observations were collected using a Python script on ten observation days, one week apart from one another, during the specified period. The script used was constructed using the Selenium library in Python, as it is considered the least intrusive to the target site (Kwon, 2023). An observation consisted of looking ahead multiple different days, using specific time intervals (in line with Abrate et al. 2019; Mohammed et al., 2019; Qiao et al., 2022), to build a panel of data of prices over time. Observing the data for multiple timepoints and over a period of time allows attempting to describe and get a better understanding of the dynamic side of the variables (Abrate & Viglia, 2016; Leoni & Nilsson, 2021).

In order to standardize the data, the price observed was for a single night in a standard double room. To enable focus on dynamic prices, the prices were observed 7, 8, 9, 10, 11, 12, 13, 15, 20, 25, 30, 45 and 60 days forward from the observation date, the simulated booking day of the imaginary customer looking for a hotel. The initial schedule consisted of 8 separate days ahead, and on 19.6.2017 five more days were added to the observations. Typically, scraping scripts, like the one implemented in this paper, can be viewed as intrusive by the targeted sites, so an initial decision was done to limit the days looked ahead. Once no complications occurred, additional dates were added to the script to gain more data and increase the reliability of the observations. The goal was to observe the impact of both time (date-of-stay close and far from the observation date) and varying demand (observing weekdays and weekends). The observation schedule is specified in the table below:

Table 2: Observation schedule

	Days ahead												
Observation date	**7**	**8**	**9**	**10**	**11**	**12**	**13**	***15***	***20***	***25***	***30***	***45***	***60***
01-Jun	08-Jun		10-Jun	11-Jun		13-Jun		16-Jun			01-Jul	16-Jul	31-Jul
08-Jun	15-Jun		17-Jun	18-Jun		20-Jun		23-Jun			08-Jul	23-Jul	07-Aug
12-Jun	19-Jun		21-Jun	22-Jun		24-Jun		27-Jun			12-Jul	27-Jul	11-Aug
19-Jun	26-Jun	27-Jun	28-Jun	29-Jun	30-Jun	01-Jul	02-Jul	04-Jul	09-Jul	14-Jul	19-Jul	03-Aug	18-Aug
26-Jun	03-Jul	04-Jul	05-Jul	06-Jul	07-Jul	08-Jul	09-Jul	11-Jul	16-Jul	21-Jul	26-Jul	10-Aug	25-Aug
03-Jul	10-Jul	11-Jul	12-Jul	13-Jul	14-Jul	15-Jul	16-Jul	18-Jul	23-Jul	28-Jul	02-Aug	17-Aug	01-Sep
10-Jul	17-Jul	18-Jul	19-Jul	20-Jul	21-Jul	22-Jul	23-Jul	25-Jul	30-Jul	04-Aug	09-Aug	24-Aug	08-Sep
17-Jul	24-Jul	25-Jul	26-Jul	27-Jul	28-Jul	29-Jul	30-Jul	01-Aug	06-Aug	11-Aug	16-Aug	31-Aug	15-Sep
24-Jul	31-Jul	01-Aug	02-Aug	03-Aug	04-Aug	05-Aug	06-Aug	08-Aug	13-Aug	18-Aug	23-Aug	07-Sep	22-Sep
31-Jul	07-Aug	08-Aug	09-Aug	10-Aug	11-Aug	12-Aug	13-Aug	15-Aug	20-Aug	25-Aug	30-Aug	14-Sep	29-Sep

The initial sample was 71 all hotels in Helsinki, but 21 were excluded from the sample at this time, because of unavailability of data. To be included in the sample a hotel had to be available on both Booking and Hotels. This excluded the 21 properties from the sample. For the remaining hotels prices were observed on 10 days, looking ahead at first for 8 days (3 times), and later for 13 days (the last 7 observation days). This resulted in prices looked for 115 booking day/date-of-stay combinations. In total for 50 hotels, on two sites, the observation of 11500 prices was attempted. The prices were not always available though, if the hotel observed was at full capacity on the desired day. Of the 11500, 1419 prices were not available, leaving the final amount of prices observed to 10081. This was deemed an adequate amount for studying the period in question. The period was also deemed long enough, so single events do not bias the results, and uniform enough (one summer season) to provide stable results. Choosing the high season as the observation point is common practice in HPMs (Arora & Mathur, 2020). Observing over a uniform season enables examining the impact of intra-week (weekdays vs. weekends) demand variations moderating the review impact on price. Below is a screenshot from Booking, captured on 19.6.2017, for a price of Radisson Blu Plaza, for a stay on 26.6.207.

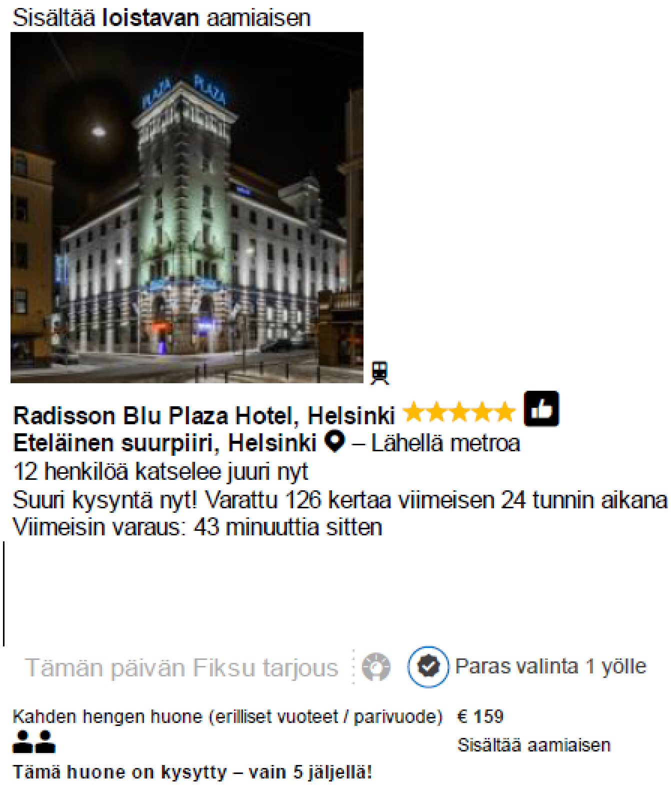

Figure 3: How a price is displayed on Booking.com

After collection, the data was put in a CSV file, with a small excerpt from the final table below, of prices observed 19.6.2017. The table has 50 rows, and 16 (first 3 observation dates) or 26 columns (the 7 last observation dates) of prices per observation date– so the whole table does not fit below. Hotel names were changed to the letter A and number for anonymity.

Table 3: An excerpt of the tabulated prices. Observed 19.6.

Hotel Code	26.6 Hotels	26.6 Booking	27.6 Hotels	27.6 Booking	continued →
A1	139	139	N/A	N/A	
A2	61	62	52	52	
A3	125	178	268	268	
A4	155	155	309	308	
A5	130	130	199	199	
A6	125	134	189	198	
A7	140	140	N/A	297	
A8	170	155	206	208	
A9	68	68	65	65	
A10	106	110	N/A	N/A	
↓ continued					

The reviews were initially scraped with another script using the Selenium library (in line with Roy, 2023) from Tripadvisor for the 50 hotels that were listed on both OTAs. In this case the review information included the posting date, the text and which hotel was commented about. Only reviews in English were considered for this study. Two random examples from the final tabulated reviews below. The index is the review's location in the data, and which is connected to a specific hotel. Date of posting is when the review was published and the text is the content extracted.

Table 4: Two review examples

Index	Date of Posting	Review
82	26-Apr-17	Recommended Cool hotel with good standard. Close to the sea where it´s very popular to walk or run. Recommend to do that. Nice pub and good breakfast. One of the bartenders might learn how to smile, but except from that; sleep well at this hotel
83	01-May-17	One night stay I was with my love in a King room. Very nice queen bed. Space is much bigger one than normally. Breakfast was great. A lot of space in toilet too. Not the best light and mirror for make-up. We will stay there again.

A further 14 properties had to be eliminated from the sample due to insufficient review activity. This meant that properties for which less than 5 reviews posted during the observation period were removed from the sample, as well as ones with less than 15 reviews in the beginning of the observation period. This left a total of 36 hotels in the sample, roughly half of the 71 hotels. The final review sample was 1923 textual comments for analysis, averaging 53,4 reviews per property analysed in this paper. The remaining hotels in the sample generated 8280 attempted price observations, 994 of these were not available, probably due to full capacity. The final observation was thus 7286 prices for 36 hotels. 202,3 prices were observed on average per hotel, and 27,6 prices were not obtained per hotel during the observation period.

EWOM generally considered as a whole, with no time aspect. The aim of this study was to measure change over time, impact on dynamic prices. In order to consider the review sentiment as a dynamic variable, its impact changing with time, they were analysed looking back from the observation date. People do not read every review when making a decision, but look 6-15 reviews back (Hospitalitynet, 2014; Ye et al., 2011). Liang et al. (2022) and Zhang et al. (2023) used a maximum of 15 most recent reviews, and this is what was chosen in this paper too.

All reviews were analysed using two sentiment analysis libraries, Textblob and VADER. These resulted in a sentiment intensity score. For each observation date, and each hotel, an average sentiment for the 15 most recent reviews was calculated with each library, establishing a moving average that changes over time as time progresses to the next observation. Analysing 15 most recent reviews enables capturing the sentiment most likely impacting the customer choices done on the observation date – and thus the dynamic price. The single observation process can be thought of from the customer viewpoint as looking back at reviews, and current prices to make a decision.

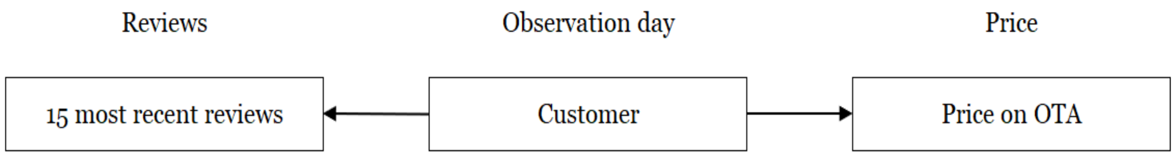

Figure 4: The simulation process

3.4 Sentiment analysis

The rise of research using unstructured data in recent years is noteworthy (Tang & Kim, 2022; Zhu, Cheng & Li, 2021; Zhang et al., 2023). Unstructured textual data provides a rich source of customer experience, emotions, sentiment, attitudes and can work in causal analysis much better than the structured, numeric data (Kwon, 2023; Tang & Kim, 2022; Xiang et al., 2017; Zhang et al., 2023). In this paper the word sentiment means the emotion expressed in the reviews. The large increase in available textual data in hospitality, combined with advancements in analysis methods can provide significant conceptual insight and new avenues for research (Kwon, 2023; Mukhopadhyay et al., 2023). In online review research (and in business intelligence) this data is commonly analysed with sentiment analysis (Geetha et al., 2017; Budianto et al., 2022 Zhu et al., 2020). Sentiment analysis (SA) is a method that enables researchers and companies to extract measures for emotions, opinions and attitudes from texts in reviews, news, Twitter and blogs for example (Kwon, 2023).

SA is a systematic analysis process for emotions that utilizes statistical methods and proprietary algorithms to identify and classify textual data, and to identify the text's polarity or valence (Hutto & Gilbert, 2014; Kwon, 2023; Majumder et al., 2022;). The process nowadays is fast, cost-efficient and reliable (Geetha et al., 2017). It is also fairly non-intrusive when implemented on Tripadvisor reviews, as they are posted anonymously.

The SA process has three main steps: acquiring the relevant data, pre-processing it and analysing the data (Xiang et al., 2015). SA can be done on several levels, ranging from the coarse negative/neutral/positive scale to a fine, continuous one, and the unit of analysis can be defined to be document-, sentence- or word-level (Hutto & Gilbert, 2014; Vuppala et al., 2022). In this paper the textual data is analysed on review-level, equating to sentence-level above. In SA papers the division of reviews into positive, neutral or negative is called the polarity-based approach, whereas discovering the intensity of the sentiment is called valence-based (Majumder et al., 2022).

Before analysis – the data used in this paper was tokenized, divided into meaningful elements called tokens – single review was made the unit to be analysed later (Xiang et al., 2017). Like in the literature (see e.g., Elbagir & Yang, 2019; Rao et al., 2020; Vuppala et al., 2022) Python library NLTK (natural language processing toolkit) was used for this for the data used in this paper.

Currently the main approaches to SA are either machine-learning or lexicon-based (Addiga & Bagui, 2022; Liu et al., 2022; Kwon, 2023). Machine-learning approach is commonly applied on document level and the lexicon-based on sentence level (Zhang et al., 2021). The machine-learning approach builds a statistical classifier based on previously labelled data the algorithm is first trained on, which can take time and resources (Zhu et al., 2020). These approaches are also found to be weaker when it comes analysing semantic relationships between words, their relative context. This weakness is remedied by using the lexicon-based approach utilizes ready-built dictionaries, that can be curated to be industry specific, of words and phrases with pre-defined sentiment score, that can be used to calculate the sentiment of the unit of analysis (Budianto et al., 2022; Liu et al., 2022; Kwon, 2023). Lexicon-based approaches are popular in hospitality and review research and identify sentiment strength on a continuous scale (Deng, Sinha & Zhao, 2017; Giatsoglou et al., 2017; Liu et al., 2022).

Textblob and VADER (short for "valence aware dictionary for sentiment reasoning") are two Python libraries that provide a normalized compound sentiment intensity score ranging from -1 (most negative sentiment in the text) to 1 (most positive sentiment), on a continuous scale, as a measure for the intensity of the sentiment expressed in the text (Addiga & Bagui, 2022; Bose, Aithal & Rot, 2020; Hutto & Gilbert, 2014; Loria, 2019).

VADER was presented to the public in Hutto & Gilbert (2014), and outperformed both human classifiers and the multiple other algorithms it was compared to, and all other seven dictionaries it was compared to (Qiao et al., 2022). Textblob utilizes naïve Bayes classifier with the Stanford natural language toolkit and applies deep learning techniques to the calculations of sentiment intensity (Giatsoglou et al., 2017; Ma, Cheng, & Hsiao, 2018; Xu & Zhao, 2022). Documentation for Textblob is provided by Loria (2019), VADER by Hutto (2021), and NLTK by NLTK (2023).

These libraries were originally developed for analysing social media comments, but are applicable for analysing the sentiment intensity expressed in the relatively short review texts (Al-Garaady & Mahyoob, 2022; Elbagir & Yang, 2019; Qiao et al., 2022). This compound score can be used for the polarity classification (negative, neutral, positive) by assigning text with a score below -0,05 as negative, text with a score over 0,05 as positive and the rest as neutral (Al-Garaady & Mahyoob, 2022; Majumder et al., 2022; Rao et al., 2020). Both libraries have been reported to be reliable and more sensitive than previous library-based approaches and human classifiers (Elbagir & Yang, 2019; Hutto & Gilbert, 2014; Kwon, 2023; Rao et al., 2020). They are used to transform unstructured, qualitative data into structured quantitative data for analysis (Zhang et al., 2021; Zhao et al., 2019). For example, Zhao et al. (2019) used the sentiment intensity degree of Textblob as an independent variable, and overall rating as a dependent variable, as did, in a similar fashion others referenced in this chapter. Majumder et al. (2022) confidently state in their study, that utilizes VADER, that the compound scores have transformed the unstructured textual data into a structured form, sentiment intensity as review valence that is considered a measure in their study.

Both libraries also consider exclamations, words, negators, phrases and word order, context, slang, conjunctions and different intensity modifiers as having different levels of intensity (e.g., "more" of something is less intensive than "most" of something) – with the result being a measure on the extent a statement is positive or negative (Addiga & Bagui, 2022; Bose et al., 2020; Rao et al., 2020; Zhao et al., 2019). As a consequence, text pre-processing like removing symbols and stop-words can be skipped in the analytical process when using these two libraries (Majumder et al., 2022; Kwon, 2023). Another consequence, as the score is a continuous variable, is that it can be used as a variable on its own, without having to use it for polarity classification before the analysis (Lai, Wang & Wang, 2021; Liang et al., 2021; Qiao et al., 2022; Roy, 2023; Xu & Zhao, 2022). This provides a richer metric on the intensity of the review, which due to its mathematical nature avoids the researcher bias, compared to methods that use human classifiers for the same task (Todri et al., 2022; Xu, 2019).

The reason for choosing to use two libraries for the analysis was to be able to compare the two different results, so the research can be cross-validated, and is reliable. This method was proposed by Kwon (2023) to increase the internal validity of the study. While it is possible to state that a higher value means a more intense sentiment, it is difficult to quantify exactly how much – and using two different libraries is a way to reduce uncertainty in the study.

3.5 Variables

In line with Illescas-Manzano et al. (2023) and Arora & Mathur (2020), this paper focuses on the review sentiment impact on dynamic price, and uses a set of other variables as control variables. The dependent variable of the study is price, and the independent variable is the review sentiment. External validity for the paper is provided by using the hedonic price model and variables the literature, based on their inclusion in prior studies (Illescas-Manzano et al., 2023; Latinopoulos, 2018).

Price is observed as a dynamic variable, in line with Abrate et al. (2019), Abrate & Viglia (2016) and Soler et al. (2019). It is considered a continuous variable, even though the observations display an intervallic nature – all prices observed were integers. To increase the reliability of the paper observations were done from two OTAs (Booking and Hotels) simultaneously. The simulation process assumes the customer will read the reviews on Tripadvisor, and then go to either Booking or Hotels to finalize the purchase. As the hotels in the sample display their prices in the two OTAs mentioned, and also do so on Tripadvisor – they have chosen to link the two (Booking, 2023_a), further motivating the choice of these two OTAs.

The independent variable is the Tripadvisor review sentiment, one review is considered as a unit of measure initially. To increase the reliability of the study, as the method used is relatively new, the sentiment intensity was quantified using two independent calculation methods, VADER and Textblob libraries in Python (see Ma et al., 2018; Li & Zhang, 2022; Kim & Han, 2022; Majumder et al., 2022 for similar methodology). The resulting score from both is a standardized continuous variable, ranging from -1 to 1, where the lower the value – the more negative the review. The individual scores were then aggregated, so an average score was calculated for the 15 most recent reviews, per property, per observation date and method. Each model specified below was then evaluated using the scores separately and the results were compared in the end. The scores show clear differences on individual reviews, but if the impact magnitude and direction acquired in the regression analysis is similar for both scores, then their usage is validated. When exploring the data from the individual scores it was noteworthy that the means differed greatly between Textblob and VADER scores (0,29 and 0,82 respectively), as did their standard deviations (0,17 and 0,9 respectively). Furthermore, it was noteworthy that the distribution of the individual reviews' VADER scores had a skew value of -3,11, which could create issues in the regressions, due to heteroskedasticity and multicollinearity, as this violates the normality assumption of regression analysis. Descriptive statistics for individual reviews below.

Table 5: Descriptive statistics for individual review sentiment

Variable	N	Mean	St.Dev	Min	Max	Skew
Textblob Score, all individual reviews	1923	0,29	0,17	-0,48	0,98	-0,44
VADER Score, all individual reviews	1923	0,82	0,39	-0,98	0,99	-3,20

An excerpt from the tabulated review data below, with the two scores calculated on the review level.

Table 6: Example reviews with sentiment scores

Index	Posting Date	Review	Textblob sentiment	Vader sentiment
1401	04/05/2017	Competition has been tightened? This hotel has been excellent before, but this time there was couple little things wrong. Bathroom sliding door was broken, bathroom was not cleaned properly (there was some previous customer's hair pins) and the breakfast waitresses were really young and didn't know what to do	-0,04	-0,76
1402	04/05/2017	Such a welcoming and great place to be! This was my third time at XXXX Hotel and it has been amazing every time! XXXX Hotel is a small luxury hotel with perfect details such as beautiful entrance hall, great towels, spacious rooms, welcoming gift/sweets, good breakfast etc. XXXX Hotel is one of the best hotels I've ever seen! I stayed in prime superior room in the renovated part of the hotel	0,49	0,99

Control variables were chosen to follow the framework of this study, which was defined as having internal and external factors that impact the price. The internal factors were the hotel type, service and property characteristics. The external factors were the site, quality signals and external market (Alderighi et al., 2022; Bigne et al., 2021; Yang & Leung, 2018). To follow the simulation process defined above, the all variables chosen were observed from Tripadvisor – which is where the customer would find the information. Dummy variables were coded so the value "1" represents the majority of the hotels in the category the dummy variable represents.

For the hotel type, two control variables were chosen: size (Binesh et al., 2021; Abrate & Viglia, 2016; Vives & Jacob, 2023) and cleanliness (Bi et al. 2019). The size variable is the number of rooms in a hotel, an interval measurement. The expected sign of the regression coefficient from the analysis below is unclear due to conflicting results in the literature (Sánchez-Lozano et al., 2021; Vives & Jacob, 2023). For cleanliness the customer evaluation of it on Tripadvisor was used. This rating is the aggregated score from one to five (at steps of 0,5 rating), an interval measure, from all reviews (until the end of the observation period) for the property. This was coded as a dummy in this paper, rating 1-4 = 0 and 4,5-5 = 1. 24 of the sampled 36 hotels had a rating of 4,5 or above. The rating is not just over the observation period and is thought to reflect reality of the hotel in this case. The expected sign of the regression coefficient for cleanliness is positive, that cleaner hotels charge more (Bi et al., 2019). Hotel style (Boutique, Romantic etc. on Tripadvisor) was considered as a variable (like in Soler et al., 2019), but this sort of categorization was deemed not useful for a small and urban market like Helsinki. For service factors, offering complimentary breakfast (Vives & Jacob, 2023) and service rating (De Olivera Santos, 2016) were chosen. Complimentary breakfast was coded as a dummy (0 = no, 1 = yes). 29 of the 36 hotels offered breakfast. Service rating on Tripadvisor is the aggregated one to five scaled (steps of 0,5) rating from all the reviews. This was coded as a dummy in this paper, rating 1-4 = 0 and 4,5-5 = 1. 19 of the 36 hotels in the sample had a service rating of 4,5 or above. It too is thought to reflect the reality of the service quality of the hotel. The property characteristic is another dummy variable, which indicates if the hotel has its own parking space (0 = no, 1 = yes) to offer to the customers (Gibbs et al, 2018; Latinopoulos 2018). 22 of the 36 hotels had their own parking facility.

For breakfast, parking and service, the expected sign of the regression coefficient is positive for all three (Arora & Mathur, 2020; Soler et al., 2019; Vives & Jacob, 2023).

As a site variable location was used. More specifically – the time (in minutes) it takes to walk from the hotel to the central railway station, as observed from Tripadvisor (Pawlicz & Napierala, 2017; Pereira & Chávez-Miranda, 2021). This variable is an interval variable. For location, a higher value means further from the railway station and these hotels should thus cost less, the expected sign of the regression coefficient is negative. For quality signals, on top of the review content defined above, Tripadvisor rating (Gibbs et al. 2018; Sánchez-Lozano et al., 2021) and chain affiliation (Sánchez-Lozano et al., 2021; Soler et al., 2019) were chosen.

Tripadvisor rating is an aggregated one to five scaled rating from all the reviews, an interval measure used as a continuous variable in this paper. Its regression coefficient is expected to be positive, that hotels with a higher rating charge more (Abrate & Viglia, 2016). It is noteworthy that on Tripadvisor the individual ratings are not included in the overall rating, all three are given independent of each other, and the service and cleanliness ratings are not used in calculating the overall rating by Tripadvisor (Qiao et a., 2022). Chain affiliation was coded as a dummy variable, 0 = no, 1 = yes. 20 of the 36 hotels in the sample were affiliated with a chain. The regression coefficient is expected to be positive for chain affiliation (Arora & Mathur, 2020). Star rating (Abrate & Viglia, 2016; Sánchez-Pérez et al., 2019; Soler et al., 2019) is a very commonly used variable in hedonic price models, but a problematic one in this context for two reasons. Firstly, there is no official star rating system in use in Finland (Visit Nordic, 2023). Secondly, the star rating is a compound score of elements already included in the model (service, parking, cleanliness for example), so it might introduce multicollinearity problems in the model (Pawlicz & Napierala, 2017).

The external market impact was controlled with three variables: popularity (Roy et al. 2022; Zeng et al, 2020), demand (Binesh et al., 2021; Abrate et al., 2019; Sánchez-Lozano et al., 2021) and time (Abrate & Viglia, 2016; Sánchez-Lozano et al., 2021; Mohammed et al., 2019; Guizzardi et al., 2022). Popularity was operationalized by counting the number of reviews posted during the observation period, and adding the initial fifteen reviews observed at the beginning of the observation. It is an interval variable, and its regression coefficient was expected to be positive, popular hotels are more in demand and are able to charge more (Abrate et al., 2019). Demand is modelled as a dummy variable, indicating whether the price observed was for a weekend (Fri/Sat), with 0 = no, 1 = yes. Weekends are expected to be more in demand and thus have a higher price (Abrate et al., 2019). The regression coefficient is expected to be positive, as prices are expected to be higher over the weekend. The final variable included in the hedonic price model is time. This is operationalized as number of days between the observation day, and the date-of-stay. Its values are the same as the days forward from the observation schedule (7, 8, 9, 10, 11, 12, 13, 15, 20, 25, 30, 45, 60) – it is considered a ratio variable in this paper, that time is continuous. As prices go up towards the approaching date-of-stay, the sign of the regression coefficient is expected to be negative, that larger values of time should have lower prices (Abrate & Viglia, 2016; Guizzardi et al., 2022).

In addition to the dependent, independent and control variables, two interaction variables are need for testing H_3 and H_4 (in line with Mariani et al., 2019). Firstly, the moderating impact of time on review sentiment's impact on price (H_3) is examined by having an interaction variable Interaction1, which is the centred time variable times centred the sentiment score (both Textblob and Vader). Centring in this context means creating a new variable by deducting the mean of the variable from the variable. Second, to test the moderating impact of high demand on review sentiment's impact on price interaction variable Interaction2 is created by multiplying the centred demand variable with the centred sentiment score (both Textblob and Vader). The baseline (when all dummy variables are zero) price is for a weekday, for a hotel that does not have a parking facility or offer complimentary breakfast. The baseline price also is for a hotel that is not chain affiliated, and has service and cleanliness ratings of 4 or below. This combination was valid for 2 of the hotels in the sample of 36.

Below a table summarizing the chosen variables, with the expected sign each regression coefficient in brackets after the name.

Table 7: Summary of the initial variables

Category	Variable	Operationalization	Measurement
Hotel type	Size (+/-)	Number of rooms	Interval
	Cleanliness (+)	Customer evaluation	Interval
Service	Breakfast (+)	Indicates if breakfast is included in the price	Dummy
	Service (+)	Customer evaluation	Interval
Property	Parking (+)	Indicates if the hotel has its own internal parking facility	Dummy
Site	Location (-)	Walking distance from the railway station	Interval
Quality signal	Online rating (+)	Customer evaluation	Interval
	Review content (+)	Sentiment intensity calculated from the review texts	Ratio
	Chain affiliation (+)	Indicates if a hotel belongs to chain	Dummy
External market	Demand (+)	Indicates if the price observed is on a Fri/Sat/Sun	Dummy
	Time (-)	The length of the booking window	Ratio
	Popularity (+)	Number of reviews at the start and posted during the observation	Interval
Interaction	Interaction1	Variable used to test H_3. Centred Review content X centred time	Ratio
Interaction	Interaction2	Variable used to test H_4. Centred Review content X centred Demand.	Ratio
Dependent	Price	Dynamic price	Ratio

After the observation period, upon inspection, several of the variables observed displayed skewed distributions. It is common practise in hedonic price modelling to solve this by log transforming the variable, and using the transformed variable instead in the regression analysis (Hu et al., 2019; Sanchez-Perez et al., 2019). Any variable which had skew > 1, was chosen to be log transformed before analysis. Price, Time, the moving average VADER score, Location and Size violated this condition. A natural logarithm was taken of these variables, and the skew value was recalculated (in line with Kim & Han, 2022; Kwon, 2023; Mariani et al., 2019). For the moving average VADER score this process increased the skew value to -

3,20, so its transform was not used. All other variables were less skewed after the transformation, so the transformation was used in the analysis. The Interaction 1 and Interaction 2 variables that used VADER score to be skewed. For these two the transform was not possible, as they included negative values. These scores were close in value to the other remaining values, log transformed and dummy variables, so a choice was made to use them nonetheless. When analysing with the interaction variables, the centred versions of time / demand were used together with the interaction variable they were used to produce. Below a table with descriptive statistics for all variables. Single star after the name means a baseline variable used in all the models. Two stars mean that the two review content variables were used in separate models in the analysis.

Table 8: Descriptive statistics of the variables

Variable	N	Mean	St.Dev	Min	Max	Skew
Price	7286	163,16	75,47	43	729	1,50
LN_Price *	7286	5,00	0,43	3,76	6,59	0,21
Time	7286	20,38	16,08	7	60	1,19
LN_Time **	7286	2,76	0,80	1,95	4,09	-0,33
LN_Time_Centred**	7286	0,00	0,80	-0,81	1,34	-0,33
Demand_Dummy **	7286	0,29	0,46	0	1	0,91
Demand_Dummy_Centred**	7286	0,00	0,46	-0,29	0,71	0,91
Review content. Textblob Score **	7286	0,29	0,06	0,12	0,46	-0,12
Review content. VADER Score **	7286	0,82	0,13	0,22	0,97	-1,80
Review content. Textblob centred **	7286	0,00	0,06	-0,18	0,17	-0,12
Review content. VADER centred **	7286	0,00	0,13	-0,59	0,15	-1,80
Popularity *	36	54,90	26,62	20	120	0,53
Parking_Dummy *	36	0,62	0,48	0	1	-0,51
Breakfast_Dummy *	36	0,82	0,38	0	1	-1,68
Location	36	12,25	8,58	1	40	1,50
LN_Location *	36	2,26	0,76	0	3,69	-0,76
Chain affiliation_Dummy *	36	0,57	0,50	0	1	-0,28
Online rating *	36	4,14	0,36	3,5	5	-0,03
Cleanliness	36	4,37	0,36	3,5	5	-0,41
Cleanliness_Dummy *	36	0,68	0,46	0	1	-0,79
Size	36	180,12	124,83	20	523	1,05
LN_Size *	36	4,92	0,81	3,00	6,26	-0,64
Service	36	4,30	0,30	4	5	0,42
Service_Dummy *	36	0,54	0,50	0	1	-0,15
Interaction1_T (LN_Time X Textblob) **	7286	0,00	0,04	-0,23	0,23	-0,03
Interaction1_V (LN_Time X VADER) **	7286	0,00	0,11	-0,79	1,63	1,27
Interaction2_T (Demand X Textblob) **	7286	0,00	0,03	-0,12	0,12	-0,06
Interaction2_V (Demand X VADER) **	7286	0,00	0,06	-0,42	0,17	-1,71

From the descriptive statistics a few observations can be made of the sample. The average price was 163,16€, clearly higher than the official 74.55€ quoted (Statistics Finland, 2018). This is most likely due to the observation period being summer, high season. The price variance was also high, with a minimum price of 43€, and a maximum of 729€ with a standard deviation of 75,47. From the sampling perspective this means that there were different kinds of hotels in the sample, making it a more representable sample. 62% of the sample had their own parking facility and 82% offered breakfast. The average walking time from hotel to the central railway station was 12,25 minutes, meaning that the hotels were situated fairly close to one another in the centre. 57% of the hotels were affiliated with a chain, and averaged 180 rooms per hotel. For the size there was considerable variation though, from 20 to 523 rooms, and a standard deviation of 124,83. And finally, the ratings for cleanliness, service as well as the overall rating all averaged over 4, meaning that customers are generally satisfied with the hotels in Helsinki.

3.6 Final model development

All in all, seven models were created for this paper. In each the dependent variable (natural logarithm of the price) was regressed against the control variables, and the dependent and interaction variables were then added for additional regressions. Due to skewness, the final dependant variable was the natural logarithm of price.

Initially, model 1 was created as a baseline, containing all the variables, except the review content, Interaction1 and interaction2. Models 2 and 3 had the review content added to them, in the form of the Textblob and VADER scores, respectively, analysed separately. Models 4 and 5 added Interaction1 (Time X Textblob and Time X VADER, respectively) to models 2 and 3. And finally models 6 and 7 added Interaction2 (Demand X Textblob and Demand X VADER, respectively) to models 2 and 3. In the equations below ε is the error term, LnPrice is a vector of the observed prices and f(x) is a linear function.

Model 1: $LnPrice_i = f(C_i) + \varepsilon_i$, where C is a vector of the control variables.

Models 2 and 3: $LnPrice_i = f(C_i, R_i) + \varepsilon_i$, where R is a vector adding review score.

Models 4 and 5: $LnPrice_i = f(C_i, R_i, I1_i) + \varepsilon_i$, I1 is a vector adding Interaction1 variable.

Models 6 and 7: $LnPrice_i = f(C_i, R_i, I2_i) + \varepsilon_i$, I2 is a vector adding Interaction2 variable.

As an example, the equation for model 2 can be written out as:

$$LnPrice_i = \beta_0 + \beta_1 * LN\_Time + \beta_2 * Demand\_Dummy + \beta_3 * Popularity + \beta_4 * Parking\_Dummy + \beta_5 * Breakfast\_Dummy + \beta_6 * LN\_Location + \beta_7 * Chain\ Affiliation\_Dummy + \beta_8 * Online\ Rating + \beta_9 * Cleanliness\_Dummy + B_{10} * LN\_Size + B_{11} * Service\_Dummy + B_{12} * Textblob\ Score + \varepsilon_i$$

Each model, and its equation was regressed and the results were obtained for analysis.

4 RESULTS

All observed variables, or their log transforms, were analysed with regression analysis. From Table 9 below which variable was used in which model can be observed. Barring a handful of exceptions, the results were statistically significant, and the coefficient signs followed the expectations to a large degree. The adjusted R2 for all models was between 37.23% and 40,67% of the variance explained, which can be considered high (Liu et al., 2019). In the table below a summary of the results, full results with t-statistics and standard errors, per model, in Appendix 1. In the table below: one star next to the coefficient has the highest significance, and three stars is the lowest, most of the selected variables were found significant. A coefficient in bold font without a star is not significant.

Table 9: Regression results for all models

Variable	Model1 Baseline	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Intercept	4,436*	4,223*	4,383*	4,554*	4,435*	4,603*	4,486*
LN_Time	-0,016**	-0,015**	-0,016**			-0,015**	-0,016**
Demand_Dummy	0,134*	0,135*	0,134*	-0,023**	-0,024**		
Popularity	0,005*	0,004*	0,005*	0,004*	0,005*	0,004*	0,005*
Parking_Dummy	0,164*	0,149*	0,160*	0,152*	0,164*	0,150*	0,161*
Breakfast_Dummy	0,109*	0,038**	0,110*	0,040**	0,110*	0,038**	0,109*
LN_Location	-0,199*	-0,179*	-0,197*	-0,179*	-0,198*	-0,179*	-0,198*
Chain affiliation_Dummy	-0,042*	-0,070*	-0,045*	-0,067*	-0,043*	-0,070*	-0,045*
Online rating	0,135*	0,090*	0,130*	0,092*	0,133*	0,090*	0,131*
Cleanliness_Dummy	0,083*	0,046*	0,075*	0,051*	0,080*	0,047*	0,075*
Service_Dummy	0,112*	0,121*	0,110*	0,118*	0,107*	0,121*	0,109*
LN_Size	**-0,016**	0,019***	**-0,013**	0,020***	**-0,012**	0,019***	**-0,013**
Textblob Score		1,165*					
VADER Score			0,080***				
LN_Time_centred				**0,000**	**-0,001**		
Textblob_centred				1,156*		1,166*	
VADER_centred					**0,073**		0,080***
Interaction1_T				-0,349*			
Interaction1_V					-0,097**		
Demand_Dummy_centred						0,135*	0,134*
Interaction2_T						**0,192**	
Interaction2_V							**0,091**
N	7286	7286	7286	7286	7286	7286	7286
F value	424,24	415,74	389,48	357,66	333,43	385,14	360,74
Adjusted R2	0,3899	0,4059	0,3902	0,3889	0,3723	0,4067	0,3909

Note: ***p<0,05, **p<0,01, *p<0,001.
VADER_centred in model 5: p=0,051.
Note: Significance F = 0 for all models

The coefficient * 100 is the percentage change in the natural logarithm of the price when the independent variable in question changes 1%. For dummy variables: (e^{β}-1) *100 (β= the coefficient) is the percentage effect of the whole category on the natural logarithm of the price, in the models in this paper (Espinet-Rius et a., 2018). Impact of the statistically significant variables from models 1, 2 and 3 below.

Table 10: Variable estimation per model

Variable	Model 1 Coefficients	Impact per variable in model 1.	Model 2 Coefficients	Impact per variable in model 2.	Model 3 Coefficients	Impact per variable in model 3.
Intercept	4,436		4,223		4,383	
LN_Time	-0,016	-1,6 %	-0,015	-1,5 %	-0,016	-1,6 %
Demand_Dummy	0,134	14,3 %	0,135	14,5 %	0,134	14,3 %
Popularity	0,005	0,5 %	0,004	0,4 %	0,005	0,5 %
Parking_Dummy	0,164	17,8 %	0,149	16,1 %	0,16	17,4 %
Breakfast_Dummy	0,109	11,5 %	0,038	3,9 %	0,11	11,6 %
LN_Location	-0,199	-19,9 %	-0,179	-17,9 %	-0,197	-19,7 %
Chain affiliation_Dummy	-0,042	-4,1 %	-0,07	-6,8 %	-0,045	-4,4 %
Online rating	0,135	13,5 %	0,09	9,0 %	0,13	13,0 %
Cleanliness_Dummy	0,083	8,7 %	0,046	4,7 %	0,075	7,8 %
LN_Size			0,019	1,9 %		
Service_Dummy	0,112	11,9 %	0,121	12,9 %	0,11	11,6 %
Textblob sentiment			1,165	116,5 %		
VADER sentiment					0,08	8,0 %

A concern in hedonic price modelling is multicollinearity, so all models were examined for it by calculating the VIF (variance inflation factor) for all independent variables in each model. The highest VIF in the models was between 3,87, and 4,16. These were far below the common threshold value of 5 (Gibbs et al., Latinopoulos, 2018; Wang et al., 2022), so multicollinearity was not an issue in any of the models.

Table 11: VIF of the independent variables in each model

Variable	Model 1 Baseline	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
LN_Time	1,02	1,02	1,02			1,03	1,03
Demand_Dummy	1,02	1,02	1,02	1,02	1,02		
Popularity	2,35	2,64	2,36	2,64	2,36	2,64	2,36
Parking_Dummy	1,33	1,35	1,38	1,35	1,38	1,35	1,38
Breakfast_Dummy	1,83	2,08	1,82	2,08	1,83	2,08	1,83
LN_Location	1,27	1,35	1,3	1,35	1,3	1,35	1,3
Chain affiliation_Dummy	2,10	2,16	2,12	1,16	2,12	2,16	2,12

Online rating	3,22	3,31	3,26	3,31	3,26	3,31	3,26
Cleanliness_Dummy	2,58	2,68	2,76	2,68	2,76	2,68	2,76
Service_Dummy	3,46	3,47	3,49	3,47	3,49	3,47	3,49
LN_Size	3,87	4,16	3,96	4,16	3,96	4,16	3,96
Textblob Score		1,84					
VADER Score			1,51				
LN_Time_centred				1,02	1,02		
Textblob_centred				1,84			
VADER_centred					1,52		
Interaction1_T				1,00			
Interaction1					1,00		
Demand_Dummy_centred						1,03	1,03
Textblob_centred						1,84	
VADER_centred							1,52
Interaction2_T						1,00	
Interaction2_V							1,00
MAX VIF / Model	3,87	4,16	3,96	4,16	3,96	4,16	3,96

In all models the most important categorical variable was the facility having its own parking facility. Of the continuous variables, location was the most important one in model 1, the review sentiment in model 2 and location again in model 3. Textblob sentiment had the biggest coefficient in models 4 and 6. Location had the biggest coefficient in models 5 and 7. All in all, the introduction of the interaction variable in models 4 through 7 did not change much the coefficients of the baseline variables, or the review content variables. In model 4 and 5 the time and demand variables were noticeably different from the other models. And location coefficient direction and significance varied from model to model. This would imply that the models were generally stable, adding to the reliability of the study.

For the time related variable, models 1, 2, 3, 6 and 7 were statistically significant ($p<0,01$), and the sign of the regression coefficient was the expected negative. This means that the longer it is from booking date to date-of-stay, the lower the price. This was in line with the literature (Abrate & Viglia, 2016; Guizzardi et al., 2022). The (centred) time variable was not statistically significant in models 4 and 5, but its impact was included in the interaction variable.

For demand, the expected sign was positive, which it was in models 1, 2, 3, 6 and 7 and the coefficients were statistically significant ($p<0,001$ in 1, 2, 3, 6 and 7, $p<0,01$ for the negative coefficients in models 4 and 5). This implies that, in Helsinki, hotels cost more over the weekend. This was further verified from the data, the average price on Fri/Sat was 182€, and averaged 155€ on the rest of the days. Popularity, also linked to demand, was also expected to show a positive sign, which it did and the coefficient was statistically significant ($p<0,001$). More popular hotels can charge more. The sign of the demand and popularity variables were mostly (not in models 4 and 5) in line with the literature (Roy et al. 2022; Zeng et al, 2020).

Parking and breakfast variables had a positive expected sign, which they were and the coefficients were statistically significant ($p<0,001$), except the breakfast variable in models 2, 4 and 6 had a lower significance ($p<0,01$). This means that adding parking and breakfast to the offering raises the price. Chain affiliation coefficient was expected to have a positive sign, which it did not. The result was statistically significant ($p<0,001$). This means that, in Helsinki, belonging to a chain means lower price. For the online overall rating, cleanliness and service the expected signs were positive, which they statistically significantly were ($p<0,001$). These were in line with the literature (Arora & Mathur, 2020; Sánchez-Lozano et al., 2021; Soler et al., 2019).

The location variable had the expected negative sign in all models, and the coefficient was statistically significant ($p<0,001$). In Helsinki, proximity to the central railway station raises prices, and hotels further away from the charge less, in line with Pereira & Chávez-Miranda (2021). The expected sign for hotel size was ambiguous in the literature (Sánchez-Lozano et al., 2021; Vives & Jacob, 2023). The results were inconclusive in this paper too. The coefficients were negative, but statistically insignificant in models 1, 3, 5 and 7 and positive and statistically significant in models 2, 4 and 6. Model 1 is considered the baseline in this paper, so this means the results for size are inconclusive.

The research question of the paper was "*How does review sentiment impact hotel dynamic prices?*". This was to be examined by accepting or rejecting 4 hypotheses, below:

H_1: Review sentiment impacts the dynamic price set by a hotel

H_2: Positive review sentiment impacts hotel price positively

H_3: Time moderates the impact of review sentiment on the dynamic price of a hotel

H_4: Demand moderates the impact of review sentiment on the dynamic price of a hotel

H_1 and H_2 were investigated by using the sentiment intensity of the written reviews, by using two Python libraries to score them. The scores were added to the baseline in models 2 and 3.

In model 2 the un-skewed Textblob score was used as a variable, and the expected sign was positive. This was the case, and the results were statistically significant ($p<0,001$). Furthermore, the magnitude of the coefficient was considerably larger than it was for the other variables. It was also statistically significant and larger than the other coefficients in models 4 and 6.

In model 3 there were concerns about the skewed VADER score data (skew = -1,8). The score utilized was found to have the expected positive sign, and was statistically significant ($p<0,05$). Albeit with lower significance and a considerably lower coefficient value (1,165 for Textblob in model 2 versus 0,08 for VADER in model 3). In model 5 the coefficient was not statistically significant (p=0,051). Both models produced results in line with the literature (Gursoy 2019; Hu et al., 2014; Liu et al., 2022; Majumder et al., 2022). **H_1** and **H_2** were **confirmed**. Review sentiment matters in pricing. In models 2, 4 and 6 review sentiment was the most important (it had the largest coefficient value) factor determining the price, while in models 3, 5 and 7, location and parking were the most important ones.

H_3 was investigated in models 4 and 5 through introducing an interaction term, which was produced by multiplying the centred LN_Time variable with the centred sentiment score. The results were statistically significant, p<0,001 for model 4 – Textblob score and time, p<0,01 for model 5 – VADER score and time. For both scores the coefficient sign was negative. This would mean that time does moderate the review impact and imply that reviews matter more the closer to consumption the booking occurs. Which is not entirely in line with the literature, and discussed more in the following chapter (Guizzardi et al., 2022; Vives, et al., 2018). Based on the results obtained **H_3** is **confirmed**.

Finally, H_4 was studied in models 6 and 7, by introducing an interaction term produced by multiplying the centred sentiment score with the centred Demand_Dummy variable. The assumption was that reviews matter less when demand is higher. The results did not support this, as they were not statistically significant in either model 6 or 7. This was not in line with the literature (Abrate, et al., 2019; Herrmann & Herrmann, 2014). **H_4** is **rejected**.

5 DISCUSSION

The results of this study have theoretical, methodological and practical implications for hospitality research and practice. Based on the hedonic price theory (HPT) literature, the assumption was that the attributes studied provide value to the customers, and their presence or absence should drive the hotels to adjust their prices. Based on the dynamic pricing literature, the assumption was that demand and time should also drive the hotels to adjust their prices further. And based on the eWOM literature the assumption was that positive review sentiment should also drive the prices up. All of these assumptions were found valid in this paper. This paper, for the first time, connected the review sentiment to price, by confirming three of the four hypotheses suggested earlier. And finally, this paper provides support for the recent methodological advances in natural language processing, supporting the idea of analysing qualitative data quantitatively.

As the models utilized HPT, they were founded in utility theory (Lancaster, 1966). Observing the different attributes present for each hotel, the customers expect to gain utility from them. They also gain utility from the positive or negative sentiment of the reviews; through the risk mitigation they provide. As the coefficient for the sentiment was considerably higher (in models 2, 4 and 6) for the sentiment, than it was for the rating – it can be stated that the sentiment provides more utility than the rating. It is noteworthy that this was not supported by models 3, 5 and 7, but the source data was more skewed in those models, making this conclusion likely. The utility expected by the customers is constantly estimated by the hotels, and used as inputs for the prices. The more they are able to do this, the closer the hotels can move towards value-based pricing, achieving a more accurate market value for their differentiated product at different times.

In a dynamic pricing context, the price observed also reflects the demand and available capacity, as those factors too are used when hotels calculate a price (Vives et al., 2019). As the other time-variant variables observed in the model are the review sentiment intensity and booking window length (which impact demand over time) – the residual, unexplained variance (1 - adjusted R^2 = ~60%) from the models contains the unobserved factors unique to each hotel and variations caused by the market conditions, for example special events.

From the models it could be seen, that in a dynamic pricing context the booking window length is a significant variable, earlier bookings can be done at a cheaper price. And while there was earlier evidence that weekends have more demand (Roma et al., 2019), it could be considered conflicting in a pricing context, as the business segment during the week is less price sensitive (Vives et al., 2019). This study found that in Helsinki the prices are higher over the weekend.

There could be variations to this finding caused by different seasons, the prices might be higher on weekdays during the winter, when the leisure demand is lower.

From the results it could be seen that location and tangible attributes matter, with parking being was the most important categorical variable in this study, and location (distance) having a large negative coefficient in the models. Hedonic price studies are usually geographically focused, as was this paper, so the importance of parking could very well just be a specific feature of the Helsinki market. This could be a sign of a large domestic segment arriving with their own vehicles, who appreciate a parking facility.

Not surprisingly, as stated by Nieto-Garcia et al. (2019) also, service was also found to be an important price determinant, as was cleanliness. Somewhat surprisingly, chain affiliation had a negative impact on the prices and the impact of hotel size was inconclusive. As chains and large properties have economies of scale on their side, maybe they are able to have a lower cost per room, and in that way are able to offer the rooms more cheaply. In all models, the review sentiment coefficient was larger than the chain affiliation variables, and the percentage impact estimated in table 10 was too. 1% change in review sentiment score impacted dynamic price more than chain affiliation. As stated in the literature, reputation is important for a service company like a hotel. The results supported this, with the overall online rating being a significant variable in all models. Review volume was found to impact price as well, so at least in the Helsinki market, this too is an important factor. Some earlier studies were opposed to each other about this (Abrate & Viglia, 2016; Gibbs et al., 2018).

The impact of different variables, that include a log transformed (and thus non-linear) dependable variable, multiple dummies and several log transformed independent variables, it can be awkward to interpret the coefficients. The impact of the variables can be illuminated by simulating the effects of each one. For this purpose, a baseline hotel scenario was created, and then each variable was changed one by one. While not all changes were possible in the real world, the simulation was easy to comprehend. In Model 2 all coefficients were statistically significant, so it was chosen for this task. In table 10 below the base case was a 100-room hotel a 10-minute walking distance from the railway station. It had 35 reviews during the observation period (and 15 before), and the Textblob score of the last 15 reviews was 0,2. Its overall rating on Tripadvisor was 4, as were the cleanliness and service ratings. It had no parking available, and no breakfast. Using the equation from model 2, the base line price for a weekday, with consumption 10 days from the booking was calculated to be 105,30€. The table below shows what happens to the price when the variables are changed one by one, with all of the other ones staying at the baseline (this method is in line with Soler et al., 2019). The dependent variable for model 2 was LN_Price.

Table 12: Simulation 1 showing the € impact of the different variables

Variable	Model 2 Coefficient	Baseline value	Action	Impact in €
Intercept	4,223	4,223		
LN_Time	-0,015	2,30 (LN of 10 days)	Book a hotel 20 days from now	-1,09 €
Demand_Dummy	0,135	0	Change from a weekday to weekend	15,22 €
Popularity	0,004	50	Add 10 reviews during the observation period	4,30 €
Parking_Dummy	0,149	0	Add parking	16,92 €
Breakfast_Dummy	0,038	0	Add breakfast	4,08 €
LN_Location	-0,179	2,30 (LN of 10 minutes)	Move hotel 5 minutes closer to railway station	13,91 €
Chain affiliation_Dummy	-0,07	0	Add hotel to a chain	-7,12 €
Online rating	0,09	4	Improve online overall rating to 4,5 or above	9,91 €
Cleanliness_Dummy	0,046	0	Improve Cleanliness rating to 4,5 or above	4,96 €
Service_Dummy	0,121	0	Improve Service rating to 4,5 or above	13,54 €
LN_Size	0,019	4,61 (LN of a 100 rooms)	Increase the size of the hotel by 1 room	0,02 €
Textblob Score	1,165	0,2	Improve Textblob score 50% from 0,2 to 0,3	13,01 €
Baseline price		105,30 €		

A second simulation was executed with a different baseline, and the variables manipulated in the opposite direction from the one in the table above. This time the baseline hotel again had a 100 rooms, and was a 10 minutes' walk from the railway station. It had 35 reviews during the observation period and the 15 before. It also had parking and complimentary breakfast, and was chain affiliated. Its online rating overall, for cleanliness and service were all a 5. The baseline price was for a weekday 10 days from now. Visible below in the table, the results when the variables were again manipulated, one by one.

Table 13: Simulation 2 showing the € impact of the different variables

Variable	Model 2 Coefficients	Baseline values	Action	Impact in €
Intercept	4,223	4,223		
LN_Time	-0,015	2,30 (LN of 10 days)	Book a hotel 20 days from now	-1,78 €
Demand_Dummy	0,135	0	Change from a weekday to weekend	24,85 €
Popularity	0,004	50	Receive 10 reviews less during the observation period	-6,75 €
Parking_Dummy	0,149	1	Remove parking	-23,81 €
Breakfast_Dummy	0,038	1	Remove breakfast	-6,42 €
LN_Location	-0,179	2,30 (LN of 10 minutes)	Move hotel a 5 minutes' walk further from the railway station	-12,04 €
Chain affiliation_Dummy	-0,07	1	Remove hotel from a chain	12,46 €
Online rating	0,09	5	Experience a dip in online overall rating to 4	-14,80 €
Cleanliness_Dummy	0,046	1	A dip in Cleanliness rating to 4	-7,74 €
Service_Dummy	0,121	1	A dip in Service rating to 4	-19,60 €
LN_Size	0,019	4,61 (LN of a 100 rooms)	Decrease the size of the hotel by 1 room	-0,04 €
Textblob Score	1,165	0,3	Dip in Textblob score 50% from 0,3 to 0,15	-27,58 €
Baseline price		171,97 €		

The main finding of this paper is the correlation between review sentiment and price, as predicted by Liu et al. (2022). Using two different methods to verify the results, both methods yielded a significant, positive correlation between the two. The intangible nature of the hotel product means that the reviews have an impact, communicating the past experience to future customers and alleviating the risk of the purchase. The better the review sentiment, the more it mitigates this risk, and the more the hotel can charge. The purchase being a subjective decision, it is no wonder that this is the case. As previous studies found that the ratings already trump hotel classifications as price determinants (Arora & Mathur, 2020: Sánchez-Pérez et al., 2019), this study found that review sentiment trumps the ratings.

As an example, a simulation was created to illustrate the power of reputation on price. In the second simulation (table 13 above), if a hotel experiences difficulties and has a simultaneous 1 rating point dip from 5 to 4 in the overall rating and a 50% dip in the sentiment of the 15 most recent reviews (measured with the Textblob method, model 2), the impact could be seen to be up to -42,38€, from the starting price of 171,97€, a 25% decrease in price. On the other hand, in simulation 1 (table 12), a 50% increase in the sentiment of the 15 most recent reviews alone (Textblob) resulted in an additional 13,01€ to a 105,3€ room, a 12% increase. This has a significant impact on the bottom line of a hotel. It is also a surprisingly large impact, as hotels can have hundreds, if not thousands of reviews and only the 15 most recent ones were looked at in this study. This paper, for the first time, examined review sentiment as a price determinant in a hedonic price model and found it significant indeed.

To examine the research question deeper, in H_3 the moderating impact of time on review sentiment impact was studied and found to be significant. The interaction variable was formed by choosing the variables of interest, deducting the mean from each and multiplying them together. The descriptives of the interaction1 and Textblob below.

Table 14: Descriptives of the first interaction variable

Variable	Min	Average	Max
LN_Time_Centred	-0,81	0	1,34
Review content. Textblob centred	-0,18	0	0,17
Interaction1_T (LN_Time X Textblob)	-0,23	0	0,23

As can be seen above, the interaction variable changes sign sometimes. The coefficient for it in models 4 and 5 had a negative sign. And its introduction made the time variable no longer significant in the models. The average of LN_Time in the data is 2,76, which translates to 15,8 days ($e^{2,76}$ = 15,8). The average of the Textblob score in the data was 0,29. In an attempt to interpret the interaction variable, 4 cases need to be specified, and the sign of each case examined.

Table 15: Interaction variable signs in different cases

		Review sentiment	
		Negative	Positive
Time	Short	+	-
	Long	-	+

In the table above a short time is 15 days or under (the average of LN_Time in days). A negative sentiment is under the average Textblob score (<0,29). From table 15, it can be interpreted that in the short run, negative reviews lower the price and positive reviews raise it. As expected. The sign of short time/negative review- combination is positive, and the coefficient of the interaction variable is negative. And as the short time/positive review combination has a negative sign, as does the coefficient, this case raises the price. Similarly, in the long run the signs in table 15 imply that positive reviews lower the price and negative ones raise it. However, the findings from models 1, 2 and 3, which were supported by literature, contradict this. This would imply further support for H_3. It implies that the review sentiment impact changes, based on the booking window length, and as the size (and thus the impact) of the time variable increases, the review sentiment weighs less in the interaction variable. The implication is that closer to consumption the review content matters more, somewhat opposed to the literature (Guizzardi et al., 2022; Herrmann & Herrmann, 2014). This could possibly be explained with reviews, being a trusted source, are used as they normally are, but the usage of other information sources is reduced due to lack of time. Alternatively, expanding on Binesh et al. (2021) and Neubert (2022): this could also be explained by strategic customers delaying their purchase, in the hopes of a better deal, who might not in the end be looking for a better price, but a better deal in general. Instead of looking for just a bargain, strategic customers could instead be looking for more value.

The last hypothesis tested was H_4, that in times of high demand the review sentiment matters less. Contradicting the literature (Ng, 2007; Abrate et al., 2019), this was rejected, meaning that the impact of the review sentiment is not moderated by demand. As H_1, H_2 and H_3 were confirmed, this means that review content is consequential even in times of high demand, and fear of missing out does not drive customers to rash decisions, and to only using ratings as heuristics.

6 LIMITATIONS AND FURTHER RESEARCH

This paper improves the understanding of the interplay between reviews and prices in hotels. It attempts to contribute by reducing the theory-practice divide in hospitality management research, through the generation of a data-driven approach studying the market and potential customers, that could be made actionable, in real time, for hotels. It also provides a way towards value-based pricing for companies, by utilizing the review content. This paper adds to the hedonic price theory research by adding a new variable – a measure for the sentiment of the reviews, and studying Helsinki as a market – which has not been done previously. This paper also adds to the eWOM literature by handling the reviews in a dynamic way, not often done in literature. The methods used in this paper can be used by practitioners to support pricing and investment decisions, as displayed by the case of the simulations. From the models it is possible to see the attribute adding value and quantify it. The methods used in this paper were novel, previously unused in both the hedonic pricing literature as well as eWOM studies on performance. Advances in research methods, in the field of natural language processing can open up new, interesting avenues for research.

There are limitations to this paper. The study was conducted in only one market, and over one season. Different markets, and the variations of the attribute importance could be studied in future research. Another limitation was that only Tripadvisor, Booking and Hotels were used as sources, future research could triangulate by using different sources. Also, only reviews in English were studied. Future research could expand by utilizing reviews in other languages. The modelling of demand was limited to only observing it on the week level. Further research could be done on the same topic matter, but using, for example demand during special events to examine the review sentiment impact under different demand conditions than were observed in this paper. Furthermore, future research could be done by including only reviews with the same rating and examining more the nuances of how the review sentiment impacts pricing.

REFERENCES

Abdullah, S., Van Cauwenberge, P., Vander Bauwhede, H. & O'Connor, P., 2022. User-generated reviews and the financial performance of restaurants. *International Journal of Contemporary Hospitality Management,* 34(10), pp. 3697-3714.

Abrate, G., Fraquelli, G. & Viglia, G., 2012. Dynamic pricing strategies: Evidence from European hotels, *International Journal of Hospitality Management*, 31, pp. 160-168.

Abrate, G. & Viglia G., 2016. Strategic and tactical price decisions in hotel revenue management, *Tourism Management*, 55, pp. 123-132.

Abrate, G., Nicolau, J.L. & Viglia, G., 2019. The impact of dynamic price variability on revenue maximization. *Tourism Management*, 74, pp. 224-233.

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

Abrate, G., Sainaghi, R. & Mauri, A.G., 2022. Dynamic pricing in Airbnb: Individual versus professional hosts. *Journal of Business Research*, 141, pp.191-199.

Addiga, A. & Bagui, S., 2022. Sentiment Analysis on Twitter Data Using Term Frequency-Inverse Document Frequency. *Journal of Computer and Communications*, 10(8), pp. 117-128.

Akhtar, N., Sun, J., Akhtar, M.N. & Chen, J., 2019. How attitude ambivalence from conflicting online hotel reviews affects consumers' behavioural responses: The moderating role of dialecticism. *Journal of Hospitality and Tourism Management*, 41, pp.28-40.

Alderighi, M., Nava, C.R., Calabrese, M., Christille, J.M. & Salvemini, C.B., 2022. Consumer perception of price fairness and dynamic pricing: evidence from Booking. com. *Journal of Business Research*, 145, pp.769-783.

Al-Garaady, J. & Mahyoob, M., 2022. Public sentiment analysis in social media on the SARS-CoV-2 vaccination using VADER lexicon polarity. *Humanities and Educational Sciences Journal*, pp. 591-609.

Alrawadieh, Z., Alrawadieh, Z. & Cetin, G., 2021. Digital transformation and revenue management: Evidence from the hotel industry. *Tourism Economics*, 27(2), pp. 328-345.

Andersson, D.E., 2010. Hotel attributes and hedonic prices: an analysis of internet-based transactions in Singapore's market for hotel rooms. *The Annals of Regional Science*, 44, pp. 229-240.

Arora, S.D. & Mathur, S., 2020. Hotel pricing at tourist destinations–A comparison across emerging and developed markets. *Tourism Management Perspectives*, 35, pp.100724.

Avlonitis, G.J. & Indounas, K.A., 2005. Pricing objectives and pricing methods in the services sector. *Journal of services marketing*, 19 (1), pp. 47–58.

Baek, H., Ahn, J. and Choi, Y., 2012. Helpfulness of online consumer reviews: Readers' objectives and review cues. International Journal of Electronic Commerce, 17(2), pp.99-126.

Barrows, C.W., Powers, T. & Reynolds, D., 2012. *Introduction to management in the hospitality industry*. New Jersey: John Wiley & Sons, pp. 406 – 424.

Bayoumi, A.E.M., Saleh, M., Atiya, A.F. & Aziz, H.A., 2013. Dynamic pricing for hotel revenue management using price multipliers. *Journal of Revenue and Pricing Management*, 12, pp.271-285.

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), pp.1-24.

Bi, J.W., Liu, Y., Fan, Z.P. & Zhang, J., 2019. Wisdom of crowds: Conducting importance-performance analysis (IPA) through online reviews. *Tourism Management*, 70, pp.460-478.

Bigne, E., Nicolau, J.L. & William, E., 2021. Advance booking across channels: The effects on dynamic pricing. *Tourism Management*, 86, pp.104341.

Binesh, F., Belarmino A., & Raab C., 2021. A meta-analysis of hotel revenue management, *Journal of Revenue and Pricing Management*, 20, pp. 546-558.

Boccali, F., Mariani, M.M., Visani, F. & Mora-Cruz, A., 2022. Innovative value-based price assessment in data-rich environments: Leveraging online review analytics through Data Envelopment Analysis to empower managers and entrepreneurs. *Technological Forecasting and Social Change*, 182, pp.121807.

Bodea, T. & Ferguson, M., 2014. *Segmentation, revenue management and pricing analytics*. London: Routledge, pp. 167 – 185.

Booking, 2023. List your hotel on Booking.com. Available at: https://join.booking.com/index.html?lang=en-gb;aid=1328032. (Accessed 16.4.2023).

Booking, 2023_a. Connecting to Tripadvisor. Available at: https://partner.booking.com/en-gb/community/new-bookingcom/connecting-tripadvisor. (Accessed 16.4.2023).

Bose, R., Aithal, P. & Roy, S., 2020. Sentiment analysis on the basis of tweeter comments of application of drugs by customary language toolkit and Textblob opinions of distinct countries. *International Journal of Emerging Trends in Engineering Research*, 8(7), pp. 3684-3696.

Bowie, D. & Buttle, F., 2004. *Hospitality marketing: An introduction*. Oxford: Elsevier. pp. 159 – 185.

Budianto, A.G., Wirjodirdjo, B., Maflahah, I. & Kurnianingtyas, D., 2022, December. Sentiment Analysis Model for KlikIndomaret Android App During Pandemic Using Vader and Transformers NLTK Library. *In 2022 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pp. 423-427. IEEE.

Cantallops, A.S. & Salvi, F., 2014. New consumer behaviour: A review of research on eWOM and hotels. *International Journal of Hospitality Management*, 36, pp.41-51.

Carroll, B., 2011. Demand Management, in Sturman, M.C., Corgel, J.B. and Verma, R. (eds.), *The Cornell school of hotel administration on hospitality: cutting edge thinking and practice*. New Jersey: John Wiley & Sons, pp. 174-191.

Castro-Lopez, A., Caso Fernández-Pacheco, L., Iglesias, V. & De la Ballina, J., 2022. Quality certifications in the TripAdvisor era. Are they outdated? *International Journal of Contemporary Hospitality Management*, 34(6), pp.2379-2398.

Chattopadhyay, M. & Mitra, S.K., 2019. Determinants of revenue per available room: Influential roles of average daily rate, demand, seasonality and yearly trend. *International Journal of Hospitality Management*, 77, pp. 573-582.

Chen, C. & Schwartz, Z., 2008. Timing Matters: Travellers' Advanced-Booking Expectations and Decisions, *Journal of Travel Research*, 47, pp. 35-42.

Chen, M. & Chen Z., 2015. Recent Developments in Dynamic Pricing Research: Multiple Products, Competition, and Limited Demand Information, *Production and Operations Management*, 24 (5), pp. 704-731.

Chen, S.X., Wang, X.K., Zhang, H.Y., Wang, J.Q. & Peng, J.J., 2021. Customer purchase forecasting for online tourism: A data-driven method with multiplex behavior data. *Tourism Management*, 87, pp.104357.

Cheng, M. & Jin, X., 2019. What do Airbnb users care about? An analysis of online review comments. *International Journal of Hospitality Management*, 76, pp.58-70.

Cirer-Costa, J.C., 2022. Qualitative revenue management in sun-and-beach hotels. *Journal of Revenue and Pricing Management*, 21(4), pp.462-469.

Cleophas, C., Kadatz, D. & Vock, S., 2017. Resilient revenue management: A literature survey of recent theoretical advances. *Journal of Revenue and Pricing Management*, 16, pp.483-498.

De Oliveira Santos, G.E., 2016. Worldwide hedonic prices of subjective characteristics of hostels. *Tourism Management*, 52, pp. 451-454.

Demirciftci, T. &d Belarmino, A., 2022. A cross-cultural study of competitive intelligence in revenue management. *Journal of Revenue and Pricing Management*, 21(6), pp.691-699.

Deng, S., Sinha, A.P. & Zhao, H., 2017. Adapting sentiment lexicons to domain-specific social media texts. *Decision Support Systems*, 94, pp. 65-76.

Dhurkari, R.K., 2023. Strategic pricing decision using the analytic hierarchy process. *Journal of Revenue and Pricing Management*, 22(1), pp.85-100.

Dickinger, A., 2011. The trustworthiness of online channels for experience-and goal-directed search tasks. *Journal of Travel Research*, 50(4), pp.378-391.

Dilme, F. & Li, F., 2012. *Revenue management without commitment: dynamic pricing and periodic fire sales*. Available at SSRN.

Dudas, G., Kovalcsik, T., Vida, G., Boros, L., & Nagy, G., 2020. Price determinants of Airbnb listing prices in Lake Balaton Touristic Region, Hungary. *European Journal of Tourism Research*, 24, pp. 1-18.

Elbagir, S. & Yang, J., 2019. Twitter sentiment analysis using natural language toolkit and VADER sentiment. *Proceedings of the international multiconference of engineers and computer scientists*, 122, pp. 16- 21.

El-Manstrly, D., Ali, F. & Line, N., 2021. Severe service failures and online vindictive word of mouth: The effect of coping strategies. *International Journal of Hospitality Management*, 95, pp.102911.

Erdem, M. & Jiang, L., 2016. An overview of hotel revenue management research and emerging key patterns in the third millennium, *Journal of Hospitality and Tourism Technology*, 7 (3), pp. 1-13.

Espinet-Rius, J.M., Fluvia-Font, M., Rigall-Torrent, R. & Oliveras-Corominas, A., 2018. Cruise tourism: A hedonic pricing approach. *European Journal of Management and Business Economics*, 27(1), pp. 101-122.

Falk, M., 2008. A hedonic price model for ski lift tickets. *Tourism Management*, 29(6), pp. 1172-1184

Fang, B., Ye, Q., Kucukusta, D. and Law, R., 2016. Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. *Tourism Management*, 52, pp.498-506.

Filieri, R., 2016. What makes an online consumer review trustworthy?, *Annals of Tourism Research*, 58, pp.46-64.

Filieri, R., Raguseo, E. & Vitari, C., 2019. What moderates the influence of extremely negative ratings? The role of review and reviewer characteristics. *International Journal of Hospitality Management*, 77, pp. 333-341.

Filieri, R., Galati, F. & Raguseo, E., 2021. The impact of service attributes and category on e-WOM helpfulness: an investigation of extremely negative and positive ratings using latent semantic analytics and regression analysis. *Computers in Human Behavior*, Vol. 114, pp. 106527.

Gao G-X. & Bi J-W., 2021. Hotel booking through online travel agency: optimal Stackelberg strategies under customer-centric payment service. *Annals of Tourism Research,* 86, pp. 103074.

Gavilan, D., Avello, M. & Martinez-Navarro, G., 2018. The influence of online ratings and reviews on hotel booking consideration. *Tourism Management*, 66, pp.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, pp. 43-54.

Gerpott, T.J. & Berends, J., 2022. Competitive pricing on online markets: a literature review. *Journal of Revenue and Pricing Management*, 21(6), pp.596-622.

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

Gibbs, C., Guttentag, D., Gretzel, U., Morton, J. & Goodwill, A., 2018. Pricing in the sharing economy: A hedonic pricing model applied to Airbnb listings. *Journal of Travel & Tourism Marketing*, 35(1), pp. 46-56.

Gretzel, U., Zarezadeh, Z., Li, Y. & Xiang, Z., 2020. The evolution of travel information search research: a perspective article. *Tourism Review*, 75(1), pp.319-323.

Guizzardi, A., Ballestra, L.V. & D'Innocenzo, E., 2022. Hotel dynamic pricing, stochastic demand and covid-19. *Annals of Tourism Research*, 97, pp.103495.

Guizzardi A., Pons F., Angelini G. & Ranieri, E., 2020. Big data from dynamic pricing: A smart approach to tourism demand forecasting. *International Journal of Forecasting*, 37(3), pp. 1049–1060.

Guizzardi, A., Mariani, M.M. & Stacchini, A., 2022. A temporal construal theory explanation of the price-quality relationship in online dynamic pricing. *Journal of Business Research*, 146, pp.32-44.

Gursoy, D., 2019. A critical review of determinants of information search behavior and utilization of online reviews in decision making process (invited paper for 'luminaries' special issue of International Journal of Hospitality Management). *International Journal of Hospitality Management*, 76, pp.53-60.

Han, W. & Bai, B., 2022. Pricing research in hospitality and tourism and marketing literature: a systematic review and research agenda. *International Journal of Contemporary Hospitality Management*, 34(5), pp. 1717-1738.

Heda, S., Mewborn, S. & Caine, S., 2017. How customers perceive a price is as important as the price itself. *Harvard Business Review*, 3, pp.5.

Hennig-Thurau, T., Gwinner, K.P., Walsh, G. & Gremler, D.D., 2004. Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet? *Journal of interactive marketing*, 18(1), pp.38-52.

Herrmann, R. & Herrmann, O., 2014. Hotel roomrates under the influence of a large event: The Oktoberfest in Munich 2012. *International Journal of Hospitality Management*, 39, pp. 21-28.

Hirschman, E., 1986. Humanistic inquiry in marketing research: philosophy, method and criteria. *Journal of marketing research*, 23(3), pp. 237-249.

Hospitalitynet.org., 2014. Hotel guests read 6-12 reviews before booking, says new TripAdvisor survey. [online] Available at: https://www.hospitalitynet.org/news/4064007.html. Accessed 14.6.2017.

Hu, F., Trivedi, R. & Teichert, T., 2022. Using hotel reviews to assess hotel frontline employees' roles and performances. *International Journal of Contemporary Hospitality Management*, 34(5), pp. 1796-1822.

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, pp. 42–53.

Hu, X., Yang, Y. & Park, S., 2019. A meta-regression on the effect of online ratings on hotel room rates. *International Journal of Contemporary Hospitality Management*, 31(12), pp.4438-4461.

Huang, G.I., Wong, I.A., Xiong, X. and Yi, K., 2022. Hotel brand equity and online reviews on social commerce intention: A cross-level identification process. *International Journal of Hospitality Management*, 105, pp.103267.

Huang, Y., Li, C., Wu, J. & Lin, Z., 2018. Online customer reviews and consumer evaluation: The role of review font. *Information & Management*, 55(4), pp. 430–440.

Hudson L. A. & Ozanne J.L., 1988. Alternative ways of seeking knowledge in consumer research. *Journal of Consumer Research*, 14 (4), pp. 508-520.

Huiyue, L., Peihan, G. & Haiwen, Y., 2022. Consistent comments and vivid comments in hotels' online information adoption: Which matters more?. *International Journal of Hospitality Management*, 107, pp.103329.

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

Hutto, C., 2021. *Vader sentiment documentation*. Available at: https://vadersentiment.readthedocs.io/en/latest/ (Accessed 16.4.2023).

Ibrahim, M. & Atiya, A., 2016. Analytical solutions to the dynamic pricing problem for time-normalized revenue, *European Journal of Operational Research*, 254, pp. 632-643.

Illescas-Manzano, M.D., Martínez-Puertas, S., Sánchez-Pérez, M. & Torres, A.M., 2023. Look before you leap: Comparison and profiles of hotel price determinants in four European markets. *International Journal of Hospitality Management*, 109, pp.103401.

Inversini, A., & Masiero, L., 2014. Selling rooms online: The use of social media and online travel agents. *International Journal of Contemporary Hospitality Management*, 26(2), pp. 272–292.

Iranmanesh, M., Ghobakhloo, M., Nilashi, M., Tseng, M.L., Yadegaridehkordi, E. & Leung, N., 2022. Applications of disruptive digital technologies in hotel industry: A systematic review. *International Journal of Hospitality Management*, 107, pp.103304.

Ivanov, S. & Zhechev, V., 2012. Hotel revenue management–a critical literature review. *Tourism: an international interdisciplinary journal*, 60(2), pp.175-197.

Ivanov, S., 2014. *Hotel revenue management: From theory to practice*. Varna: Zangador, pp. 8-163.

Ivanov S., Del Chiappa G. & Heyes A., 2021. The research-practice gap in hotel revenue management: insights from Italy. *International Journal of Hospitality Management* 95, pp. 102924.

Janssen, E., 2021. Changes in pre-and post-pandemic pricing decision factors: An overview of South Africa's luxury accommodation sector. *Research in Hospitality Management*, 11(1), pp.37-44.

Jiang, Q., Liu, S., Tang, H. & Liu, Y.Q., 2021. Research on the online reviews semantic network and perceptual dimension of Short – Rent B&B – text mining based on Tujia and Airbnb. *Resource Development and Market*, 38 (2), pp. 239-248.

Kaemingk, D., 2022. Online review statistics to know in 2022. Available at: https://www.qualtrics.com/blog/online-review-stats/ (Accessed 16.4.2023).

Kahneman, D. & Tversky, A., 1979. Prospect theory: An analysis of decision under risk. *Econometrica*, 47, pp. 263-292.

Kim, J.M. & Han, J., 2022. Impact of the length of stay at hotels on online reviews. *International Journal of Contemporary Hospitality Management*, 34(2), pp. 1249-1269.

Kim, J.Y. & Hwang, J., 2022. Who is an evangelist? Food tourists' positive and negative eWOM behavior. *International Journal of Contemporary Hospitality Management*, 34(2), pp.555-577.

Kim, J., Kim, S.S., Jhang, J., Lee, J. & Koo, C., 2023. Effects of price sorting display on extreme option choice aversion: The role of ease of comparison in multiple option displays. *Tourism Management*, 97, pp.104741.

Kimes, S. E., 1989. The basics of yield management. *The Cornell Hotel and Restaurant Administration Quarterly*, 30(3), pp. 14–19.

Kimes, S., 2011. Revenue Management for Enhanced Profitability, in Sturman, M.C., Corgel, J.B. and Verma, R. (eds.), *The Cornell school of hotel administration on hospitality: cutting edge thinking and practice*. New Jersey: John Wiley & Sons, pp. 192-206.

Klein, R., Koch, S., Steinhardt, C. & Strauss, A., 2019. A review of revenue management: Recent generalizations and advances in industry applications, *European Journal of Operational Research*, June, pp. 1-16.

Kool, S., Westerlaken, R. & Suleri, J., 2022. Dynamic pricing and perceived fairness: a case study at a hotel on the West Frisian Island of Vlieland, The Netherlands. *Research in Hospitality Management*, 12(2), pp.143-150.

Kwon, W., 2023. Reading customers' minds through textual big data: Challenges, practical guidelines, and proposals. *International Journal of Hospitality Management*, 111, pp.103473.

Labsomboonsiri, S., Mathews, S., Luck, E. & Mulcahy, R., 2022. Recognize me or pay me? How rewards for online restaurant reviews impact continuity: A cross-country investigation. *International Journal of Hospitality Management*, 106, pp.103289.

Lancaster, K.J., 1966. A new approach to consumer theory. *Journal of political economy*, 74(2), pp.132-157.

Lai, X., Wang, F. & Wang, X., 2021. Asymmetric relationship between customer sentiment and online hotel ratings: the moderating effects of review characteristics. *International Journal of Contemporary Hospitality Managemen*t, 33(6), pp.2137-2156.

Latinopoulos, D., 2018. Using a spatial hedonic analysis to evaluate the effect of sea view on hotel prices. *Tourism Management*, 65, pp. 87-99.

Lee S., 2011. *Study of demand models and price optimization performance*. Doctoral Thesis, Georgia Tech Theses and Dissertations, School of Industrial and Systems Engineering. Available at: https://smartech.gatech.edu/bitstream/handle/1853/42914/lee_seonah_201112_phd.pdf (Accessed: 11.04.2023).

Lee, S., Garrow, L.A., Higbie, J.A., Keskinocak, P. & Koushik, D., 2011. Do you really know who your customers are? A study of US retail hotel demand. *Journal of Revenue and Pricing Management*, 10, pp.73-86.

Leoni, V. & Nilsson, W., 2021. Dynamic pricing and revenues of Airbnb listings: Estimating heterogeneous causal effects. *International Journal of Hospitality Management*, 95, pp.102914.

Li, X. & Zhang, Y., 2022. A prediction framework on the helpfulness of reviews for processing Chinese and English hotel reviews. *International Journal of Contemporary Hospitality Management*, 34(12), pp. 4415--4437.

Liang, S., Zhang, X., Li, C., Li, H. & Yu, X., 2021. Tit for tat: understanding the responding behavior of property hosts on peer-to-peer rental platforms. *International Journal of Contemporary Hospitality Management,* 33(3), pp, 1105-1126.

Liang, S., Ye, Q., Zhang, X., Law, R. & Gong, C., 2022. Motivation behind review provision in online travel communities: do hometowner contributions matter? *International Journal of Contemporary Hospitality Management*, 34(5), pp. 1692-1716.

Ling, L., Guo, X. & Yang, C., 2014. Opening the online marketplace: An examination of hotel pricing and travel agency on-line distribution of rooms. *Tourism management*, 45, pp.234-243.

Ling, L., Dong, Y., Guo, X. & Liang, L., 2015. Availability management of hotel rooms under cooperation with online travel agencies. *International Journal of Hospitality Management*, 50, pp.145-152.

Litvin, S.W., Goldsmith, R.E. & Pan, B., 2008. Electronic word-of-mouth in hospitality and tourism management. *Tourism management*. 29, pp. 458-468.

Liu, J., Yu, Y., Mehraliyev, F., Hu, S. & Chen, J., 2022. What affects the online ratings of restaurant consumers: a research perspective on text-mining big data analysis. *International Journal of Contemporary Hospitality Management*, 34(10), pp. 3607-3633.

Lopes, A.I., Dens, N., De Pelsmacker, P. & Malthouse, E.C., 2023. Managerial response strategies to eWOM: A framework and research agenda for webcare. *Tourism Management*, 98, pp.104739.

Loria, S., 2019. *Textblob documentation*. Available at: https://textblob.readthedocs.io/en/dev/ (Accessed 16.4.2023).

Lozano, J., Rey-Maquieira, J., & Sastre, F., 2021. An integrated analysis of tourism seasonality in prices and quantities, with an application to the Spanish hotel industry. *Journal of Travel Research*, 60(7), pp. 1581–1597.

Ma, E., Cheng, M. & Hsiao, A., 2018. Sentiment analysis–a review and agenda for future research in hospitality contexts. *International Journal of Contemporary Hospitality Management*, 30(11), pp. 3287-3308.

Majumder, M.G., Gupta, S.D. & Paul, J., 2022. Perceived usefulness of online customer reviews: A review mining approach using machine learning & exploratory data analysis. *Journal of Business Research*, 150, pp.147-164.

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

Mariani, M.M. & Visani, F., 2019. Embedding eWOM into efficiency DEA modelling: An application to the hospitality sector. *International Journal of Hospitality Management*, 80, pp.1-12.

Mariani, M., 2020. Big data and analytics in tourism and hospitality: a perspective article. *Tourism Review*, 75(1), pp.299-303.

Matsuoka, K., 2022. Effects of revenue management on perceived value, customer satisfaction, and customer loyalty. *Journal of Business Research*, 148, pp.131-148.

Mauri, A. G., & Minazzi, B., 2013. Web reviews influence on expectations and purchasing intentions of hotel potential customers. *International Journal of Hospitality Management*, 34, pp. 99–107.

Mehraliyev, F., Chan, I.C.C. & Kirilenko, A.P., 2022. Sentiment analysis in hospitality and tourism: a thematic and methodological review. *International Journal of Contemporary Hospitality Management*, 34(1), pp. 46-77.

Melis, G. & Piga, C.A., 2016. Are all online hotel prices created dynamic? An empirical assessment, *MPRA Paper*, 75896, pp. 1-37.

Millauer, T., &Vellekoop, M., 2019. Artificial intelligence in today's hotel revenue management: opportunities and risks, *Research in Hospitality Management*, 9(2), pp. 121–124.

Mohammed, I., Guillet, B.D. & Law, R., 2019. Last-minute hotel-booking and frequency of dynamic price adjustments of hotel rooms in a cosmopolitan tourism city. *Journal of Hospitality and Tourism Management*, 41, pp.12-18.

Mukhopadhyay, S., Pandey, R. & Rishi, B., 2023. Electronic word of mouth (eWOM) research–a comparative bibliometric analysis and future research insight. *Journal of Hospitality and Tourism Insights*, 6(2), pp.404-424.

Nagle, T.T. & Holden, R.K., 1995. *The Strategy and Tactics of Pricing*, Prentice-Hall, Englewood Cliffs, NJ.

Neirotti, P., Raguseo, E. & Paolucci, E., 2016. Are customers' reviews creating value in the hospitality industry? Exploring the moderating effects of market positioning. *International Journal of Information Management*, 36(6), pp.1133-1143.

Neubert, M., 2022. A systematic literature review of dynamic pricing strategies. *International Business Research*, 15(4), pp. 1-17.

Ng, F., Rouse, P. & Harrison, J., 2017. Classifying revenue management: A taxonomy to assess business practice. *Decision Sciences*, 48(3), pp. 489-522.

Ng, I.C., 2007. *The pricing and revenue management of services: A strategic approach*. London: Routledge, pp. 24-197.

Nicolau, J.L. & Sharma, A., 2019. To Ban or not to ban rate parity, that is the question… or not? *International Journal of Hospitality Management*, 77, pp. 523-527.

Nie, R.X., Tian, Z.P., Wang, J.Q. & Chin, K.S., 2020. Hotel selection driven by online textual reviews: Applying a semantic partitioned sentiment dictionary and evidence theory. *International Journal of Hospitality Management*, 88, pp. 102495.

Nieto-Garcia, M., Munoz-Gallego, P. & Gonzalez-Benito, O., 2017. Tourists' willingness to pay for an accommodation: The effect of eWOM and internal reference price. *International journal of hospitality management*. 62, pp. 67-77.

Nieto-Garcia, M., Resce, G., Ishizaka, A., Occhiocupo, N. & Viglia, G., 2019. The dimensions of hotel customer ratings that boost RevPAR. *International Journal of Hospitality Management*, 77, pp. 583-592.

NLTK, 2023. *NLTK documentation*. Available at: https://www.nltk.org/ (Accessed 16.4.2023).

Noone, B.M., 2016. Pricing for hotel revenue management: Evolution in an era of price transparency. *Journal of Revenue and Pricing Management*, 15, pp.264-269.

O'Connor, P., 2020. Online tourism and hospitality distribution: a perspective article. *Tourism Review*, 75(1), pp.290-293.

Pawlicz, A. & Napierala, T., 2017. The determinants of hotel room rates: an analysis of the hotel industry in Warsaw Poland, *International Journal of Contemporary Hospitality Management*, 29 (1), pp.571-588.

Petricek, M., Chalupa, S. & Chadt, K., 2020. Identification of consumer behaviour based on price elasticity: A case study of the Prague market of accommodation services. *Sustainability*, 12(22), pp. 1-14.

Petricek, M., Chalupa, S. & Melas, D., 2021. Model of price optimization as a part of hotel revenue management—stochastic approach. *Mathematics*, 9(13), pp. 15- 52.

Petricek, M., Chalupa, Š. & Levickova, V., 2022. Comparison of expected marginal revenue models in the hospitality industry. *Journal of Revenue and Pricing Management*, 21(3), pp.299-305.

Phillips, P., Zigan, K., Silva, M.M.S. & Schegg, R., 2015. The interactive effects of online reviews on the determinants of Swiss hotel performance: A neural network analysis. *Tourism Management*, 50, pp.130-141.

Phillips, P., Barnes, S., Zigan, K., & Schegg, R., 2017. Understanding the impact of online reviews on hotel performance: an empirical analysis. *Journal of Travel Research*, 56(2), pp. 235-249.

Prakash, D. & Spann, M., 2022. Dynamic pricing and reference price effects. *Journal of Business Research*, 152, pp.300-314.

Priester A., Robbert T. & Roth S., 2020. A special price just for you: effects of personalised dynamic pricing on consumer fairness perceptions. *Journal of Revenue and Pricing Management,* 19(2), pp. 99–112.

Qiao, T., Shan, W., Zhang, M. & Wei, Z., 2022. More than words: Understanding how valence and content affect review value. *International Journal of Hospitality Management*, 105, pp.103274.

Raad, J., Sharma, A. & Nicolau, J.L., 2023. Performance effects of innovation in two-sided markets: The paradigmatic case of OTAs. *Tourism Management*, 94, pp.104637.

Raguseo, E., & Vitari, C., 2017. The Effect of Brand on the Impact of e-WOM on Hotels' Financial Performance. *International Journal of Electronic Commerce*, 21 (2), pp.249 - 269.

Rao, S., Monica, N., Nikhila, P., Tejasri, T. and Maram, B., 2020. Positivity Calculation using Vader Sentiment Analyser. *International Journal of Academic Engineering Research*, 4, pp.13-17.

Rigall-I-Torrent, R. & Fluvia, M., 2011. Managing tourism products and destinations embedding public good components: a hedonic approach. *Tourism Management*, 32(2), pp.244-255.

Rita, P., Ramos, R., Borges-Tiago, M.T. & Rodrigues, D., 2022. Impact of the rating system on sentiment and tone of voice: A Booking. com and TripAdvisor comparison study. *International Journal of Hospitality Management*, 104, pp.103245.

Roma, P., Panniello, U. & Nigro, G., 2019. Sharing economy and incumbents' pricing strategy: The impact of Airbnb on the hospitality industry. *International Journal of Production Economics*, 214, pp.17-29.

Ropero, M. A., 2011. Dynamic pricing policies of hotel establishments in an online travel agency. *Tourism economics*, 17(5), pp. 1087-1102.

Rosen, S., 1974. Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of political economy*, 82(1), pp.34-55.

Roy, G., Datta, B., Mukherjee, S. & Shrivastava, A.K., 2022. Systematic review of eWOM literature in emerging economy using ACI framework. *International Journal of Emerging Markets*, ahead-of-print.

Roy, G., 2023. Travellers' online review on hotel performance–Analysing facts with the Theory of Lodging and sentiment analysis. *International Journal of Hospitality Management*, 111, pp.103459.

Sanchez-Lozano G., Pereira L.N. & Chávez-Miranda E., 2021. Big data hedonic pricing: econometric insights into room rates' determinants by hotel category. *Tourism Management*, 85, pp. 104308.

Sanchez-Pérez, M., Illescas-Manzano, M.D. & Martínez-Puertas, S., 2019. Modelling hotel room pricing: A multi-country analysis. *International Journal of Hospitality Management*, 79, pp.89-99.

Sann, R., Lai, P.C. & Chen, C.T., 2021. Review papers on eWOM: prospects for hospitality industry. *Anatolia*, 32(2), pp.177-206.

Saunders, M., Lewis, P. & Thornhill, A., 2009. *Research methods for business students*. Pearson education.

Shadiqurrachman, S., Ridwan A.Y. & Kusuma P.G.A., 2019. Online Travel Agency Channel Pricing Policy based on Dynamic Pricing Model to Maximize Sales Profit Using Nonlinear Integer Programming Approach, *Advances in Intelligent Systems Research*, 171, pp. 351-358.

Soler, I. P., Gemar, G., Correia, M. B. & Serra, F., 2019. Algarve hotel price determinants: A hedonic pricing model. *Tourism Management*, 70, pp. 311–321.

Statistics Finland, 2018. *Nights spent in all establishments*. Available at: https://www.stat.fi/til/matk/2017/matk_2017_2018-04-19_tau_001_en.html. (Accessed 16.4.2023).

Talluri, K.T., Van Ryzin, G. & Van Ryzin, G., 2004. *The theory and practice of revenue management (Vol. 1)*. Boston: Kluwer Academic Publishers, pp. 175-236.

Talon-Ballestero, P., Nieto-Garcia, M. & Gonzalez-Serrano, L., 2022. The wheel of dynamic pricing: Towards open pricing and one to one pricing in hotel revenue management. *International Journal of Hospitality Management*, 102, pp.103184.

Tang, M. & Kim, H.S., 2022. An exploratory study of electronic word-of-mouth focused on casino hotels in Las Vegas and Macao. *Information*, 13(3), pp.135-157.

Tavares, V.C., Tavares, F. & Santos, E., 2022. The Value of Farmland and Its Determinants—The Current State of the Art. *Land*, 11(11), pp.1908 – 1930.

Tietokeskus, 2022. Helsinki facts and figures 2022. Available at: https://www.hel.fi/static/hel2/tietokeskus/julkaisut/pdf/22_06_15_Helsinki_facts_and_figures_2022.pdf. (Accessed 16.4.2023).

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), pp.118-140.

Tran, L.T.T., 2020. Online reviews and purchase intention: A cosmopolitanism perspective. *Tourism Management Perspectives*, 35, pp.100722.

Tripadvisor, 2023. About Tripadvisor. Available at: https://tripadvisor.mediaroom.com/us-about-us (accessed 16.4.2023)

Viglia G & Abrate G., 2019. Revenue and yield management: a perspective article. *Tourism Review,* 75(1), pp. 294–298.

Visit Nordic, 2023. What do the hotel stars mean. Available at: https://www.visitnordic.com/en/article/what-do-the-hotel-stars-mean. (Accessed 16.4.2023).

Vives, A., Jacob, M. & Payeras, M., 2018. Revenue management and price optimization techniques in the hotel sector: A critical literature review. *Tourism economics*, 24(6), pp.720-752.

Vives, A., Jacob, M. & Aguilo, E., 2019. Online hotel demand model and own-price elasticities: An empirical application in a mature resort destination. *Tourism Economics*, 25(5), pp. 670-694.

Vives, A. & Jacob, M., 2023. Sources of price elasticity of demand variability among Spanish resort hotels: a managerial insight. *Journal of Hospitality and Tourism Technology*, 14(2), pp. 137-153.

Vempala, S., Singe, S., Vasa, S. & Bandla, K., 2022. Twitter Sentiment Analysis Using NLTK and Machine Learning. *International Journal for Research in Applied Science & Engineering Technology*, 10(7), pp. 1494-1499.

Vuppala, S., Singa, S., Vasa, S. & Bandla, K., 2022. Twitter Sentiment Analysis Using NLTK and Machine Learning. *International Journal for Research in Applied Science & Engineering Technology*, 10(7), pp. 1494-1499.

Wamsler, J., Natter, M. & Algesheimer, R., 2022. Transitioning to dynamic prices: Should pricing authority remain with the company or be delegated to the service employees instead? *Journal of Business Research*, 139, pp.1476-1488.

Wang, E.Y., Fong, L.H.N. & Law, R., 2022. Detecting fake hospitality reviews through the interplay of emotional cues, cognitive cues and review valence. *International Journal of Contemporary Hospitality Management*, 34(1), pp.184-200.

Wang, X., Tang, L.R. & Kim, E., 2019. More than words: Do emotional content and linguistic style matching matter on restaurant review helpfulness? *International Journal of Hospitality Management*, 77, pp. 438-447.

Wang, Y., Kim, J. & Kim, J., 2021. The financial impact of online customer reviews in the restaurant industry: A moderating effect of brand equity. *International Journal of Hospitality Management*, 95, pp.102895.

Wilk, V., Lambert, C. and Meek, S., 2022. Online brand detraction in an online opinion platform. *International Journal of Hospitality Management*, 107, pp.103345.

Williams, N.L., Ferdinand, N. & Bustard, J., 2020. From WOM to aWOM–the evolution of unpaid influence: a perspective article. *Tourism Review*, 75(1), pp.314-318.

Xiang, Z., Schwartz, Z., Gerdes Jr, 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, pp.120-130.

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, pp.51-65.

Xie, K.L., Chen, C. & Wu, S., 2016. Online consumer review factors affecting offline hotel popularity: evidence from Tripadvisor. *Journal of Travel & Tourism Marketing*, 33(2), pp.211-223.

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), pp.673-683.

Xu, X., Zhang, L., Baker, T., Harrington, R.J. & Marlowe, B., 2019. Drivers of degree of sophistication in hotel revenue management decision support systems. *International Journal of Hospitality Management*, 79, pp.123-139.

Xu, X., 2019. Examining the relevance of online customer textual reviews on hotels' product and service attributes. *Journal of Hospitality & Tourism Research*, 43(1), pp.141-163.

Xu, X. & Zhao, Y., 2022. Examining the influence of linguistic characteristics of online managerial response on return customers' change in satisfaction with hotels. *International Journal of Hospitality Management*, 102, pp.103146.

Yang, Y. & Leung, X. Y., 2018. A better last-minute hotel deal via app? Cross-channel price disparities between Hotel Tonight and OTAs. *Tourism Management*. 68, pp. 198-209.

Yang, Y., Park, S. & Hu, X., 2018. Electronic word of mouth and hotel performance: A meta-analysis. *Tourism management*, 67, pp.248-260.

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

Ye, Q., Law, R., Gu, B. & Chen, W., 2011. The influence of user-generated content on traveller behaviour: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings. *Computers in Human behavior*, 27(2), pp. 634-639.

Ye, X., Fu, Y.K., Wang, H. & Zhou, J., 2023. Information asymmetry evaluation in hotel E-commerce market: Dynamics and pricing strategy under pandemic. *Information Processing & Management*, 60(1), pp. 103117.

Yen, C. & Tang, C., 2019. The effects of hotel attribute performance on electronic word-of-mouth (eWOM) behaviours. *International Journal of Hospitality Management*, 76, pp.9-18.

Yeoman, I., 2016. The history of revenue and pricing management–15 years and more. *Journal of Revenue and Pricing Management*, 15, pp.185-196.

Zaki, K., 2022. Implementing dynamic revenue management in hotels during Covid-19: value stream and wavelet coherence perspectives. *International Journal of Contemporary Hospitality Management*, 34(5), pp. 1768-1795.

Zeng, G., Cao, X., Lin, Z. & Xiao, S.H., 2020. When online reviews meet virtual reality: Effects on consumer hotel booking. *Annals of Tourism Research*, 81, pp.102860.

Zhang, C., Xu, Z., Gou, X. & Chen, S., 2021. An online reviews-driven method for the prioritization of improvements in hotel services. *Tourism Management*, 87, pp.104382.

Zhang, D. & Weatherford, L., 2017. Dynamic pricing for network revenue management: A new approach and application in the hotel industry. *INFORMS Journal on Computing*, 29(1), pp.18-35.

Zhang, X., Grisolia, J.M. & Lane, T.D., 2023. Zero price effect on hotel demand: Evidence from a discrete choice experiment. *Tourism Management*, 96, pp.104692.

Zhang, X., Zhang, X., Liang, S., Yang, Y. & Law, R., 2023. Infusing new insights: How do review novelty and inconsistency shape the usefulness of online travel reviews? *Tourism Management*, 96, pp.104703.

Zhang, Y., Gao, J., Bilgihan, A. & Lorenz, M., 2023. A holistic assessment of eWOM management effectiveness with agent-based modelling. *International Journal of Contemporary Hospitality Management*, 35(3), pp.785-827.

Zhang, Z., Liang, S., Li, H., & Zhang, Z., 2019. Booking now or later: Do online peer reviews matter? International *Journal of Hospitality Management*, 77, pp. 147–158.

Zhang, Z., Ye, Q. & Law, R., 2011. Determinants of hotel room price: An exploration of travellers' hierarchy of accommodation needs. *International Journal of Contemporary Hospitality Management*, 23(7), pp. 972-981.

Zhao, X., Wang, L., Guo, X. & Law, R., 2015. The influence of online reviews to online hotel booking intentions. *International Journal of Contemporary Hospitality Management*, 27(6), pp. 1343-1364.

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, pp.111-121.

Zhu, J., Cheng, M. & Li, Z., 2021. Peer-to-peer accommodation experience and guest actual recommendations: A novel mixed-method approach. *Tourism Management Perspectives*, 38, pp. 100816.

Zhu, L., Lin, Y. & Cheng, M., 2020. Sentiment and guest satisfaction with peer-to-peer accommodation: when are online ratings more trustworthy? *International Journal of Hospitality Management*, 86, pp. 102369.

APPENDICES

Appendix 1: Regression results for each model individually

Appendix 1.1 - Model 1: baseline model output

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,6252
R Square	0,3908
Adjusted R Square	0,3899
Standard Error	0,3341
Observations	7286

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	11	520,8192	47,3472	424,2	0
Residual	7274	811,8138	0,1116		
Total	7285	1332,6330			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	4,4364	0,0861	51,5323	0,0000	4,2676	4,6051
LN_Time	-0,0157	0,0049	-3,1870	0,0014	-0,0254	-0,0061
Demand_Dummy	0,1342	0,0087	15,3499	0,0000	0,1171	0,1513
Popularity	0,0046	0,0002	20,5792	0,0000	0,0042	0,0051
Parking_Dummy	0,1640	0,0093	17,6030	0,0000	0,1457	0,1823
Breakfast_Dummy	0,1087	0,0138	7,8587	0,0000	0,0816	0,1358
LN_Location	-0,1993	0,0058	-34,5807	0,0000	-0,2106	-0,1880
Chain affiliation_Dummy	-0,0424	0,0115	-3,7023	0,0002	-0,0649	-0,0200
Online rating	0,1351	0,0197	6,8525	0,0000	0,0965	0,1738
Cleanliness_Dummy	0,0830	0,0135	6,1381	0,0000	0,0565	0,1095
LN_Size	-0,0162	0,0095	-1,7129	0,0868	-0,0348	0,0023
Service_Dummy	0,1124	0,0146	7,6968	0,0000	0,0838	0,1410

Appendix 1.2 - Model 2: Textblob score impact model output

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,6379
R Square	0,4069
Adjusted R Square	0,4059
Standard Error	0,3297
Observations	7286

ANOVA

	df	SS	MS	F	Significance F
Regression	12	542,1973	45,1831	415,74	0
Residual	7273	790,4357	0,1087		
Total	7285	1332,6330			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	4,223	0,0863	48,9268	0,0000	4,0536	4,3920
LN_Time	-0,015	0,0049	-3,1367	0,0017	-0,0249	-0,0057
Demand_Dummy	0,135	0,0086	15,6470	0,0000	0,1181	0,1519
Popularity	0,004	0,0002	15,0121	0,0000	0,0031	0,0040
Parking_Dummy	0,149	0,0093	16,1485	0,0000	0,1313	0,1676
Breakfast_Dummy	0,038	0,0146	2,6107	0,0091	0,0095	0,0665
LN_Location	-0,179	0,0059	-30,3886	0,0000	-0,1901	-0,1671
Chain affiliation_Dummy	-0,070	0,0115	-6,0656	0,0000	-0,0921	-0,0471
Online rating	0,090	0,0197	4,5373	0,0000	0,0508	0,1282
Cleanliness_Dummy	0,046	0,0136	3,4163	0,0006	0,0198	0,0731
LN_Size	0,019	0,0097	2,0014	0,0454	0,0004	0,0383
Service_Dummy	0,121	0,0144	8,4064	0,0000	0,0930	0,1495
Textblob score	1,165	0,0831	14,0252	0,0000	1,0020	1,3276

Appendix 1.3 - Model 3: VADER score impact model output

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,6255
R Square	0,3912
Adjusted R Square	0,3902
Standard Error	0,3340
Observations	7286

ANOVA

	df	SS	MS	F	Significance F
Regression	12	521,34450	43,44537	389,47699	0
Residual	7273	811,28853	0,11155		
Total	7285	1332,63303			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	4,38315	0,08949	48,97820	0,0000	4,20772	4,55858
LN_Time	-0,01587	0,00494	-3,21159	0,0013	-0,02555	-0,00618
Demand_Dummy	0,13438	0,00874	15,37404	0,0000	0,11725	0,15152
Popularity	0,00460	0,00023	20,37743	0,0000	0,00416	0,00505
Parking_Dummy	0,16034	0,00947	16,93669	0,0000	0,14178	0,17890
Breakfast_Dummy	0,10962	0,01384	7,92232	0,0000	0,08250	0,13674
LN_Location	-0,19742	0,00583	-33,87777	0,0000	-0,20885	-0,18600
Chain affiliation_Dummy	-0,04492	0,01151	-3,90143	0,0001	-0,06748	-0,02235
Online rating	0,13046	0,01983	6,57797	0,0000	0,09158	0,16934
Cleanliness_Dummy	0,07515	0,01399	5,37004	0,0000	0,04772	0,10258

LN_Size	-0,01318	0,00956	-1,37838	0,1681	-0,03193	0,00556
Service_Dummy	0,10966	0,01465	7,48294	0,0000	0,08093	0,13838
VADER Score	0,08025	0,03698	2,17008	0,0300	0,00776	0,15274

Appendix 1.4 - Model 4: Interaction of Time and Textblob output

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,6245
R Square	0,3900
Adjusted R Square	0,3889
Standard Error	0,3339
Observations	7286

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	13	518,5081	39,8852	357,6639	0
Residual	7272	810,9442	0,1115		
Total	7285	1329,4523			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	4,5537	0,0857	53,1202	0,0000	4,3857	4,7218
LN_Time_centred	-0,0002	0,0049	-0,0431	0,9656	-0,0098	0,0094
Demand_Dummy	-0,0231	0,0079	-2,9099	0,0036	-0,0386	-0,0075
Popularity	0,0036	0,0002	15,1140	0,0000	0,0031	0,0041
Parking_Dummy	0,1521	0,0094	16,2252	0,0000	0,1337	0,1704
Breakfast_Dummy	0,0398	0,0147	2,7011	0,0069	0,0109	0,0687
LN_Location	-0,1793	0,0060	-30,1259	0,0000	-0,1910	-0,1677
Chain aff_Dummy	-0,0672	0,0116	-5,7801	0,0000	-0,0899	-0,0444
Review rating	0,0917	0,0200	4,5862	0,0000	0,0525	0,1308
Cleanliness_Dummy	0,0507	0,0138	3,6788	0,0002	0,0237	0,0777
Service_Dummy	0,1184	0,0146	8,1022	0,0000	0,0897	0,1470
LN_Size	0,0199	0,0098	2,0268	0,0427	0,0007	0,0391
Textblob_centred	1,1565	0,0841	13,7461	0,0000	0,9915	1,3214
Interact1_T	-0,3486	0,0765	-4,5563	0,0000	-0,4986	-0,1986

Appendix 1.5 - Model 5: Interaction of Time and VADER output

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,6111
R Square	0,3735
Adjusted R Square	0,3723
Standard Error	0,3384
Observations	7286

ANOVA

	df	SS	MS	F	Significance F
Regression	13	496,4983	38,1922	333,4320	0
Residual	7272	832,9539	0,1145		
Total	7285	1329,4523			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	4,4351	0,0865	51,2528	0,0000	4,2655	4,6048
LN_Time_centred	-0,0012	0,0050	-0,2360	0,8134	-0,0109	0,0086
Demand_Dummy	-0,0236	0,0080	-2,9430	0,0033	-0,0394	-0,0079
Popularity	0,0047	0,0002	20,3464	0,0000	0,0042	0,0051
Parking_Dummy	0,1637	0,0096	17,0671	0,0000	0,1449	0,1825
Breakfast_Dummy	0,1102	0,0140	7,8601	0,0000	0,0827	0,1377
LN_Location	-0,1984	0,0059	-33,5874	0,0000	-0,2099	-0,1868
Chain aff_Dummy	-0,0429	0,0117	-3,6761	0,0002	-0,0658	-0,0200
Review rating	0,1328	0,0201	6,6064	0,0000	0,0934	0,1722
Cleanliness_Dummy	0,0796	0,0142	5,6102	0,0000	0,0518	0,1073
Service_Dummy	0,1074	0,0148	7,2325	0,0000	0,0783	0,1365
LN_Size	-0,0124	0,0097	-1,2769	0,2017	-0,0314	0,0066
VADER_centred	0,0730	0,0375	1,9478	0,0515	-0,0005	0,1465
Interact1_V	-0,0972	0,0351	-2,7706	0,0056	-0,1660	-0,0284

Appendix 1.6 - Model 6: Interaction of Demand and Textblob output

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,6386
R Square	0,4078
Adjusted R Square	0,4067
Standard Error	0,3290
Observations	7286

ANOVA

	df	SS	MS	F	Significance F
Regression	13	542,0958	41,6997	385,1369	0
Residual	7272	787,3565	0,1083		
Total	7285	1329,4523			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	4,6032	0,0854	53,9226	0,0000	4,4359	4,7706
LN_Time	-0,0152	0,0049	-3,1284	0,0018	-0,0248	-0,0057

Demand_Dummy_Centred	0,1349	0,0086	15,6635	0,0000	0,1180	0,1518
Popularity	0,0035	0,0002	15,0285	0,0000	0,0031	0,0040
Parking_Dummy	0,1499	0,0092	16,2282	0,0000	0,1318	0,1680
Breakfast_Dummy	0,0378	0,0145	2,6024	0,0093	0,0093	0,0663
LN_Location	-0,1787	0,0059	-30,4575	0,0000	-0,1902	-0,1672
Chain aff_Dummy	-0,0695	0,0115	-6,0708	0,0000	-0,0920	-0,0471
Review rating	0,0897	0,0197	4,5551	0,0000	0,0511	0,1283
Cleanliness_Dummy	0,0467	0,0136	3,4381	0,0006	0,0201	0,0733
Service_Dummy	0,1209	0,0144	8,3988	0,0000	0,0927	0,1491
LN_Size	0,0195	0,0097	2,0185	0,0436	0,0006	0,0384
Textblob_centred	1,1660	0,0829	14,0659	0,0000	1,0035	1,3285
Interact2_T	0,1916	0,1362	1,4065	0,1596	-0,0755	0,4587

Appendix 1.7 - Model 7: Interaction of Demand and VADER output

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,6261
R Square	0,3921
Adjusted R Square	0,3910
Standard Error	0,3334
Observations	7286

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	13	521,2210	40,0939	360,7420	0
Residual	7272	808,2313	0,1111		
Total	7285	1329,4523			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	4,4861	0,0862	52,0562	0,0000	4,3171	4,6550
LN_Time	-0,0159	0,0049	-3,2204	0,0013	-0,0256	-0,0062
Demand_Dummy_centred	0,1345	0,0087	15,4113	0,0000	0,1174	0,1516
Popularity	0,0046	0,0002	20,3906	0,0000	0,0042	0,0050
Parking_Dummy	0,1606	0,0094	16,9980	0,0000	0,1421	0,1792
Breakfast_Dummy	0,1093	0,0138	7,9101	0,0000	0,0822	0,1363
LN_Location	-0,1975	0,0058	-33,9469	0,0000	-0,2089	-0,1861
Chain aff_Dummy	-0,0451	0,0115	-3,9219	0,0001	-0,0676	-0,0225
Review rating	0,1308	0,0198	6,6049	0,0000	0,0920	0,1696
Cleanliness_Dummy	0,0752	0,0140	5,3836	0,0000	0,0478	0,1026
Service_Dummy	0,1093	0,0146	7,4748	0,0000	0,0807	0,1380
LN_Size	-0,0129	0,0095	-1,3543	0,1757	-0,0316	0,0058
VADER_centred	0,0802	0,0369	2,1732	0,0298	0,0079	0,1526
Interact2_V	0,0909	0,0658	1,3822	0,1669	-0,0380	0,2199