Predictors of Active Loyalty: The case of Hotel Group X

Ana Vera Nascimento Prada

Master in Marketing

Supervisor:
Dr. João Guerreiro, Assistant Professor, Department of Marketing, Operation and General Management,
ISCTE-IUL

November, 2021

(This page was intentionally left blank)

Resumo

Os programas de fidelização são, atualmente, considerados padrões da indústria no sector hoteleiro. Tais programas visam encorajar compras recorrentes, recompensar clientes fiéis, assim como atrair novos, aumentar as taxas de retenção e a quota de mercado, e melhorar a recolha de informação sobre os clientes. No entanto, a simples participação num programa de fidelização não implica uma lealdade ativa. Este projeto in-company procura identificar os clientes leais ativos do Grupo Hoteleiro X, fornecendo à empresa informações sobre quem são agora esses hóspedes e quais poderão vir a sê-lo no futuro, permitindo-lhes conceber estratégias de marketing apropriadas.

Neste estudo foi utilizada a metodologia CRISP-DM com o principal objetivo de descobrir as variáveis que mais influenciam a troca de pontos por recompensas, e que, por sua vez, se traduzem em lealdade ativa. Foram utilizados dois modelos: C&RT e a Regressão Logística. De acordo com os resultados do C&RT, as reservas feitas no website da empresa são as preditoras mais importantes de recompensas redimidas, seguidos de estadias na região do Algarve e estadias em hotéis urbanos. Já no modelo de Regressão Logística foi possível concluir que os clientes corporate são muito significativos nesta previsão. Para além disso, pudemos concluir que todos os canais diretos de marcação de estadias são, também, preditores.

Os nossos resultados podem, assim, ajudar a melhorar a direção prática da empresa, que lida com um grande volume de dados, podendo estes serem eventualmente integrados nos modelos construídos neste estudo, de forma a gerar novos conhecimentos sobre os consumidores.

Palavras-chave: Indústria Hoteleira, Lealdade Ativa, Programas de Fidelização, CRM, Big Data, C&RT, Regressão Logística

JEL Sistema de Classificação: Marketing (M31); Métodos de Classificação (C38); Métodos de Previsão e Predição (C53)

Abstract

Loyalty programs are now considered industry standards in the hotel sector. Such programs aim to encourage repeat purchases, attract new customers, reward loyal ones, increase retention rates and market share, and collect customer information. Nonetheless, simple participation in a loyalty program does not imply active loyalty. This in-company project seeks to identify Hotel Group X's active loyal customers and provide the company with insights into who these guests are today and who may become one in the future, allowing them to design appropriate marketing strategies.

The CRISP-DM methodology was employed in this study, and its data mining goals were to uncover the most important predictors of reward redemptions, which translate into active loyalty. Two predictive models were used in this study – C&RT and Logistic Regression. According to the C&RT model, reservations made on the company's website are the best predictor of reward redemptions, followed by stays in the Algarve region and city hotels. The Logistic Regression model suggests that there is a significant predictive power for the corporate customers, followed by all the direct booking channels.

Our results can help enhance the practical direction for hotel managers who deal with vast volumes of data that can be further integrated into the model built in this study to generate novel insights on consumers.

Keywords: Hospitality Industry, Active Loyalty, Loyalty Programs, CRM, CRISP-DM, Big Data, C&RT, Logistic Regression

JEL Classification System: Marketing (M31); Classification Methods (C38); Forecasting and Prediction Methods (C53)

(This page was intentionally left blank)

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AUC: Area Under Curve

BD: Big Data

C&RT: Classification and Regression Trees

CRISP-DM: Cross-industry Process for Data Mining

CRM: Customer Relationship Management

FN: False Negative

FP: False Positive

F&B: Food and Beverage

LP: Loyalty Program

LR: Logistic Regression

OTA: Online Travel Agencies

RM: Relationship Marketing

TN: True Negative

TP: True Positive

UGC: User-generated Content

UNWTO: World Tourism Organization

WOM: Word of Mouth

(This page was intentionally left blank)

Introduction

Tourism has been steadily rising for decades, and it is now one of the world's fastest-growing industries. Tourism has a business volume that rivals or exceeds oil, food, and autos. As a result, it has grown to be one of the most critical participants in international trade and one of the primary sources of income for many developing countries. This expansion is accompanied by increased destination diversification and competitiveness (UNWTO, 2021).

In 2017, international visitor arrivals increased by 7.0%, the most significant rate since the global economic crisis of 2009. In actual terms (adjusted for currency rate variations and inflation), international tourism receipts climbed 4.9% to US$ 1,340 billion in the same year. According to the UNWTO, international tourist arrivals (overnight visitors) grew by 6.0% to 1.4 billion in 2018, well above the global economy's 3.7% growth. In 2019, international tourist visits globally increased to 1.5 billion. It saw continued robust growth, albeit at a slower pace than the exceptional rates of 2017 and 2018 (UNWTO, 2020). When direct, indirect, and induced consequences are considered, travel and tourism account for 10.4% of global GDP and one out of every ten employment (WTTC, 2020). According to the latest figures from the World Tourism Organization, worldwide tourism had its worst year on record in 2020 due to the COVID-19 pandemic, with international arrivals falling by 74% (UNWTO, 2020). Due to a historic drop in demand and extensive travel restrictions, destinations worldwide will welcome 1 billion fewer international arrivals than the previous year (UNWTO, 2020).

Tourism is also a critical activity for Portugal's economic growth and employment. Over the last nine years, the country has recorded an average annual growth rate of 7.2% of overnight stays, resulting in a jump from 37 million overnight stays in 2010 to 70 million overnight stays in 2019 - the most significant increase in history. Similarly, receipts increased at a 10.3% average annual rate, totaling 18.4 billion euros in 2019. In the same year, the tourism industry accounted for 17.1% of the national GDP (Turismo de Portugal, 2021).

According to estimates, the number of guests arriving in all forms of tourist accommodation, in 2019, totaled 29.5 million and the number of overnight stays stood at 77.8 million (INE, 2020, p. 5). Overnight stays in tourist accommodations totaled 70.2 million and hotels recorded 58 million overnight stays. The internal market provided 26.1 million overnight stays, accounting for 33.6% of total overnight stays, and increased by 5.9%. Overnight stays for foreign markets rose by 3.5% to 51.7 million (66.4 % of the total) (INE, 2020, p. 27). The number of non-resident tourists arriving in Portugal in 2019 should have reached 24.6 million, representing a 7.9% increase over the previous year. Spain remained the leading inbound market (25.5%). Tourists from the United Kingdom (15.4% of all visitors) increased by 7.6%. The arrivals of French tourists (a share of 12.6%) increased by 2.1%, causing this country to lose some of its representativeness. On the other hand, the German market (7.9%) showed no variation, whereas the Brazilian market (5.5 % of the total) increased by 13.9%. The emphasis outside the European Union was on the 23.2% increase in the number of tourists arriving from the United States (INE, 2020, p. 5).

The number of non-resident tourist arrivals in Portugal will have reached 6.5 million in 2020, representing a 73.7% decrease from 2019. Because this year was marked by the COVID-19 pandemic, the presented results reflect the particularly harmful effects on the tourism sector, with significant reductions (INE, 2021, p. 5).

Businesses are concerned about customer optimism as the world joins forces to contain the present COVID-19 pandemic. The two pillars of client loyalty, trust and confidence, are being tested. While we all hope that this is a one-time occurrence, anxiety is high, and people are afraid. This global problem is genuinely about critical consumer moments. By putting the consumers' needs first, managers can position their company's brand to take the lead (Main et al., 2020). While no one knows what these new realities will look like, businesses will actively work toward stabilization and recovery, which in the travel and hospitality industries includes ensuring that a brand's most valuable customers return as soon as possible. Loyalty programs can be very effective in recovering loyalists (Glassoff et al., 2020).

Loyalty programs are one of the most effective ways for companies to collect information about their customers and, when done correctly, to improve customer satisfaction and generate new purchases (Segel et al., 2013). Hotels have made significant investments in loyalty over the last decade, and more than 90% of businesses participate in some form of loyalty program. Between 2012 and 2014, the number of memberships in the United States alone increased by 26.7%, and there are now 3.3 billion memberships, implying that each household has 29 memberships (Woolan, 2017). According to Accenture research (2017), loyalty program members generate between 12% and 18% more revenue than non-members. At the same time, according to the Boston Consulting Group, some companies generate up to 60% of their revenue from loyalty program members (Bolden et al., 2014).

Nonetheless, mere enrollment in a loyalty program does not indicate active loyalty (Jennings et al., 2014), and it is argued that investments done by hotels do not generate as much value as they could. Part of the reason for this is that loyalty programs do not prevent brand switching. Hotel loyalty members have a surprisingly low affinity for their preferred brand and spend up to 50% of their wallets on non-preferred brands (Weissenberg et al., 2013).

Furthermore, traditional loyalty schemes require a rethink. These loyalty programs are no longer sufficient to satisfy more demanding customers, who differ in terms of expectations and responses to triggers. According to Deloitte research, 53% of customers do not even redeem their points on a regular basis. They would rather be rewarded based on their preferences. As a result, members place a premium on personalization and relevance regarding what they expect from the program. Given the importance of customer data in allowing hotels to provide more personalized experiences, Deloitte research revealed that nearly one-third of customers are willing to exchange their data for better benefits (Fenech & Perkins, 2017).

Fortunately, in recent years, the hospitality industry has begun to capitalize on the vast amounts of data available, improving its analytics capabilities in order to better anticipate and meet customer needs and preferences (Bhattacharjee et al., 2017). Artificial Intelligence, Machine Learning, and the Internet of Things are flourishing and, when combined, have the potential to create personalized moments that matter to the customer, displacing the one-size-fits-all mentality that no longer works in the hospitality industry (Weissenberg & Langford, 2018). Revenue management was the first area to use advanced analytics at scale, implementing dynamic pricing, which is now an industry standard (Bhattacharjee et al., 2017). In addition, hotels are increasingly using data to better understand their guests and aid in real-time decision-making. Through the combination of client feedback and transactional data (e.g., room bookings, food, and drink transactions, CRM activity), hotels can uncover the profile of the guests (their age bracket, spending power and preferences, where they come from, and the reason for their stay). This information can be used to predict future stays and improve retention (Higgins, 2020).

Furthermore, businesses will be able to create unique offers and experiences in real-time. A hotel, for example, is using next-product-to-buy algorithms that analyze historical data to determine if, say, a customer who is traveling with a spouse is likely to enjoy early-morning coffee. Then, using cell phone location, a buy-one-get-one-free offer is delivered just as the customer walks by the hotel coffee shop in the morning (Bhattacharjee et al., 2017). Similarly, if a guest is interested in music, they will receive a push notification linked to discounted tickets for a jazz show downtown (Weissenberg & Langford, 2018).

The issue, however, is in the real-time operationalization and collection of relevant data from digital channels or through employees. These examples demonstrate the future of data-driven customer-centric value creation (Weissenberg & Langford, 2018). There will be significant opportunities for improving the guest experience and developing a competitive advantage as soon as hospitality companies recognize and commit to the potential of these advanced analytics.

This project is being developed within the context of Hotel Group X and its loyalty program. The group acknowledged many flaws and weaknesses in their strategy, which resulted in a low redemption rate. The redemption rate for a loyalty program is the percentage of points awarded and redeemed for prizes (Burnett, 2020). Members of loyalty programs earn points, usually based on the volume, value, and frequency of their spending. The member redeems the points accumulated and receives various rewards, such as free flights, cash back, money off, or gifts (Smith & Sparks, 2009). This project will address the problem described above, by studying Hotel Group X's active loyal consumers (customers who redeem rewards).

Hotel Group X was founded in 1972 in Madeira Island, Portugal. It expanded to mainland Portugal in the early 1990s, Africa in the late 1990s, and South America in the final year of the twentieth century. Hotel Group X expanded into other European countries and, later, the United States at the start of the second decade of the twenty-first century. It is typically ranked among the top 25 and top 100 hotel groups in Europe and globally, respectively. It owns approximately 13000 rooms in around 100 hotels, across 15 countries, divided into four categories: luxury, hotels and resorts, lifestyle, and historical monuments. It has always been a 100% private group owned by a single person. In 2019, the group's annual turnover was around € 400 million, with an EBITDA of € 175 million.

Hotel Group X has a loyalty program that allows guests to receive various benefits and advantages before, during, and after their stays. Immediately after registering, the guest gets a 10% discount on bookings made through the Hotel Group X's website. There are five tiers: Guest, Elite, Elite Plus, Honor, and Corporate. Based on their tier level, guests receive different discounts on accommodation, food and beverage (F&B), and spa services. Moreover, upper-tier level guests receive preferential treatment such as complimentary room service, priority check-in, early check-in, late check-out, and room upgrade discounts. In addition, Hotel Group X provides its guests with discounts and special conditions at various partners, including stores, museums, and other activities.

Depending on their tier level, guests earn between 10 and 15 points per euro spent in bookings at the company's website or via other direct channels, as well as in F&B. However, if they choose to book through online travel agencies (OTAs), they will receive a fixed amount of 250 points. These points can be redeemed for partly free nights (>2500 points plus cash) or full free nights (>18000 points). In addition, the group has special offers and promotions from time to time in which guests can exchange smaller amounts of points (from 1000 to 4000 points plus cash) for stays in specific hotel units.

Currently, 1.4 million guests are members of the loyalty program. However, only 900,000 of these guests made a reservation between 2018 and 2020, with 50 thousand booking multiple times. Furthermore, the redemption rate was minimal during 2018 and 2019 – 1% and 2%, respectively.

Therefore, this project intends to identify active loyal customers (customers who redeem rewards), discover patterns and common characteristics among them and develop models capable of predicting who can become one in the future. At the same time, this project aims to analyze members of the program who do not redeem rewards and how they differentiate from those who do. The overarching purpose is to provide the company with meaningful insights that could aid in designing marketing strategies and decision-making.

1. Literature Review

This project will use data mining and predictive techniques to analyze active loyal customers from Hotel Group X's loyalty program. Loyalty programs are Customer Relationship Management (CRM) tools that aim to reward loyal customers, increase repeat purchases, and gather customer data, among other things. The role of technology in hotel CRM is critical. With the introduction of Big Data, it is now possible to manage such data to achieve goals by transforming information into knowledge. Big Data analytics, which are powerful tools in the hospitality industry, were also used in this project. Taking all the above into consideration, this chapter will cover de concepts of Customer Loyalty, Loyalty Programs, CRM, and Big Data.

1.1. Loyalty

1.1.1. Customer Loyalty

Customer loyalty is defined as repeat visitation or purchase behavior where the customer also holds an emotional commitment or a positive attitude towards the service provider (Petrick, 2004; Shoemaker and Lewis, 1999). Oliver (1999) once described brand loyalty as "a deeply held commitment to rebuy or repatronize a preferred product/service consistently in the future, thereby causing repetitive same-brand or same brand-set purchasing, despite situational influences and marketing efforts having the potential to switching behavior" (Oliver, 1999, p. 34). The worth of loyal clients is significant because they visit more frequently and make more purchases than non-loyal customers do (Yoo & Bai, 2013). Furthermore, it is believed that a slight increase in loyal consumers can result in a considerable rise in profitability (Reichheld, 1993; Reichheld and Sasser, 1990) and that attracting a new client is six times more expensive than retaining an existing one (Petrick, 2004).

Authors divide customer loyalty into three categories: behavioral, attitudinal, and composite loyalty. Behavioral loyalty refers to regular, repetitive purchasing behavior (Bowen & Chen, 2001). Ultimately, it assesses the likelihood of acquiring/using the service and future purchasing intent (Yoo & Bai, 2013). The emotional and psychological attachment to a brand is shown in the attitudinal dimension (Bowen & Chen, 2001). Trust, emotional attachment or commitment, and switching costs are essential elements of attitudinal loyalty (Baloglu, 2002). The composite loyalty argues that neither behavioral nor attitudinal loyalty alone are sufficient to describe loyalty. Instead, assessing brand loyalty requires understanding both aspects (Dick & Basu, 1994; Chaudhuri & Holbrook, 2001).

Researchers have recognized customer satisfaction, loyalty programs, switching costs, service quality, and commitment crucial to customer loyalty (Kandampully & Suhartanto, 2000; Lee et al., 2001; Hu et al., 2010). Customer satisfaction and commitment will positively impact repurchase (Garbarino & Johnson, 1999). Trust also positively impacts guests' loyalty (O'Mahony et al., 2013) because it is interrelated with commitment. They are both essential to any successful business relationship (Morgan and Hunt, 1994; Bowen and Shoemaker, 2003). Since service quality influences customer satisfaction, it will positively affect loyalty as well (Demirci Orel & Kara, 2014). Loyalty programs were predicted to directly influence customer loyalty (Tanford et al., 2011; Hu et al., 2010), along with switching costs (Baloglu, 2002).

Loyal consumers foster strong relationships with companies, resulting in beneficial results such as repurchase intentions, increased share of wallet, word of mouth (WOM), and lower acquisition costs. All of these should contribute to increased profitability for the company (Zeithaml, 2000; Chi & Gursoy, 2009; Dalci et al., 2010; Evanschitzky et al., 2012). In addition, loyal customers are willing to pay more to the same seller, implying a larger share of wallet (Palmatier et al., 2006). In the hospitality industry, the value of a guest for one hotel, as a proportion of his overall value at all other hotels, is referred to share of wallet (Xie & Chen, 2015). Moreover, enterprises with long-term clientele can charge higher prices for their products (Reichheld & Sasser, 1990), and loyal customers are prepared to pay a fair premium (Palmatier et al., 2007). Furthermore, customer loyalty can present itself in two ways: recommending and patronizing the service provider to others (Lam et al., 2004). WOM is a critical outcome of customer-firm relationships (Brown et al., 2007; Reichheld, 2003), and it is seen as a reliable and trustworthy source of information that will positively affect the decision-making of other potential guests (Litvin et al., 2008; Park & Lee, 2009; Zhang et al., 2010).

1.1.2. Loyalty Programs

While academics spent decades researching the mechanics underlying consumer loyalty, practitioners developed and implemented techniques to grow a loyal customer base. For example, American Airlines, which developed its first loyalty program in 1981, intending to increase repeat purchases (McCall & Voorhees, 2010). Loyalty programs (LP) seem to have become popular in the hospitality industry (Xie & Chen, 2013). LPs are critical Customer Relationship Management tools (Kumar & Reinartz, 2018) aiming at boosting repeat purchases, acquiring new consumers, rewarding loyal ones, increasing retention rates and market share, and collecting customer information (Xie & Chen, 2015; Hendler et al., 2021). However, whether these programs benefit the organizations is still controversial (Xie et al., 2015). It is stated that loyalty programs may fail to grasp customer types, expectations, and behaviors, such as deal-seekers joining the program and making one-time purchases (Pesonen et al., 2019). Although major hotels such as Hilton, Marriott, and InterContinental have over ten million loyalty program members, a large portion of them are deal-seekers who participate in many hotel programs and are not necessarily committed to a specific one (Xie & Chen, 2015; Hendler et al., 2021). As a result, hotels must understand how consumers perceive and use loyalty programs to attract the right ones and create long-term, mutually beneficial relationships with them (Hendler & LaTour, 2008; Liu-Thompkins & Tam, 2013).

Researchers believe that adequately designed loyalty programs can increase repeat purchases, willingness to pay a price premium, share of wallet, and advocacy (Sharp & Sharp, 1997; Verhoef, 2003; Keh & Lee, 2006; Leenheer, 2007). However, research suggests that a loyalty program's effectiveness depends upon three factors - the structure of the loyalty program, the structure of the rewards, and consumer fit with the loyalty program (McCall & Voorhees, 2010). Therefore, hotel managers should understand what types of structure and rewards the different types of customers prefer (Pesonen et al., 2019).

1.1.2.1. Tiers

Usually, hotels use three different tiers - regular, middle, and elite. Drèze & Nunes (2009) believe that three-tier programs bring the most satisfaction to all members, especially those in the elite tier, which are provided with a sense of status. Moreover, tiers can help segment customers and provide differentiated rewards (Rigby & Ledingham, 2004). In addition, customers' buying behavior, namely their frequency and magnitude of consumption, may be influenced by their transition between tiers (McCall & Voorhees, 2010).

1.1.2.2. Rewards

Hotels reward members with benefits that can be categorized into tangible or intangible. The former benefits are comprised of economic benefits such as immediate discounts (Pesonen et al., 2019), free hotel stays, tickets (McCall & Voorhees, 2010) or prizes (Lee et al., 2015). On the other hand, intangible rewards include special treatment and privileges such as members-only newsletters (Pesonen et al., 2019), personalized recognition, and special services (Lee et al., 2015). Research suggests that intangible rewards build more affective commitment (Lee et al., 2015) and generate more attitudinal loyalty (Pesonen et al., 2019). Furthermore, regarding the reward’s timing, customers recognize more value in immediate rewards than in delayed ones, which are reported to work better only when customers are already satisfied (Lee et al., 2015).

The rewards component is thought to influence loyalty via point pressure and the rewarded behavior effect. Customers' switching costs and future orientation can result in the formation of the points pressure effect. Customers do not want to purchase from another brand when they are getting close to a reward. As a result, people boost their purchasing and spending in order to obtain points to exchange for rewards (Kopalle et al., 2012). The rewarded behavior effect, on the other hand, occurs after the redemption. Because of behavioral learning (Rothschild & Gaidis, 1981), the belief in a gain or a good deal, a sense of appreciation (Smith and Sparks, 2009), or an elevated sense of status (Drèze & Nunes, 2009), once a consumer redeems the rewards, the frequency and volume of purchases will grow.

1.1.2.3. Customer Fit

A major success aspect in any loyalty program is whether the consumer sees and identifies himself with the benefits of membership and the extent to which his needs and purchasing behavior align with the program (Pesonen et al., 2019). Furthermore, loyalty programs that are a good fit may foster a sense of community among members (Pesonen et al., 2019; McCall & Voorhees, 2010). Customers' involvement, perceived fit, and status perceptions will influence their perception of the program (McCall & Voorhees, 2010). As a result, the program must provide the "best" incentives to the "best" customers (Pesonen et al., 2019).

1.1.3. Active Loyalty

Loyalty programs are intended to keep clients continuously engaged (Xie et al., 2015). Point accumulation thresholds and tier transaction options are used by companies to encourage customers to make repeat purchases. However, as previously said, and despite hotel efforts, membership does not always result in active loyalty. Active loyalty is defined by the customer's premeditated behavior and willingness to extend the relationship with the hotel (Xie & Chen, 2014; Xie et al., 2015). Therefore, active loyalty can be conceptualized as “customers’ active engagement with any functions of a loyalty program, such as making reservations through the program, point accumulation, and/or reward redemptions” (Xie et al., 2015). Moreover, active loyalty is also linked to future usage, enhanced word of mouth, and, ultimately, to profitability. Additionally, it is believed to be a more efficient customer loyalty measure (Xie & Chen, 2015).

1.2. CRM

When airlines pioneered its frequent flyer program to reward loyal customers in the 1960s, they also pioneered the concept of Relationship Marketing (RM) (Xiong et al., 2014). RM focuses on directing a company's marketing actions toward establishing and maintaining relationships with customers through the creation of mutual benefits (Kim & Cha, 2002). This notion was then converted into CRM, in subsequent research, incorporating a wide range of perspectives ranging from a technological solution (Payne & Frow, 2005) to a customer-centric process (Garrido-Moreno & Padilla-Melendez, 2011) and a management philosophy (Vaeztehrani et al., 2015).

According to Boulding et al. (2005), the field of CRM has begun to converge around a common meaning. Frow and Payne (2009) argued that “CRM is a cross-functional strategic approach concerned with creating improved shareholder value through the development of appropriate relationships with key customers and customer segments. It typically involves identifying the right business and customer strategies, the acquisition and diffusion of customer knowledge, deciding appropriate segment granularity, managing the co-creation of customer value, developing integrated channel strategies, and the wise use of data and technology solutions to create superior customer experiences”.

1.2.1. CRM Implementation

Different authors have proposed different models for successfully implementing CRM (Rahimi, 2017). CRM is a multidimensional construct comprised of four fundamental behavioral components, according to Sin et al. (2005): key customer factors, CRM organization, knowledge management, and technology-based CRM. Customer orientation, however, has been discovered to be more comprehensive than key customer factors (Mohammad, 2013), especially in the hotel industry, where the notion has been employed to increase connections between customers and hotel organizations (Wu & Lu, 2012). Sigala (2005), on the other hand, suggested that an integrated managerial approach involved three domains, namely information and communications technology (ICT), relationship, and knowledge management. According to Chen and Popovich (2003) and Mendoza et al. (2007), CRM is a combination of people, processes, and technology, and a holistic approach between these three components is essential for a successful CRM implementation.

The three views have a lot in common when it comes to successfully deploying CRM. Overall, such implementation requires aligning people with new strategies and processes, organizational readiness and collaboration with employees, as well as a shift in the direction of organizations' operations from product-centric to customer-centric. Technology is critical in implementing the CRM strategy and business re-design. It involves collecting and analyzing data on customer patterns, interpreting customer behavior, creating a 360-degree view of customers, developing predictive models, responding with timely and effective customized communications, and delivering product and service value to individual customers (Rahimi & Gunlu, 2016).

1.2.2 Types of CRM (Operational, Collaborative and Analytical)

Operational CRM systems help to automate CRM processes and improve the efficiency of customer-facing operations. These include customer service and support systems like call centers, sales force automation (e.g., point of sale systems), and marketing automation (Khodakarami & Chan, 2014).

Collaborative CRM involves the processes that enable communication and interaction with the customers. It manages the channels and the touchpoints such as websites, e-mail, customer portals, and video conferencing. Therefore, it helps the continuous acquisition and generation of customer knowledge (Khodakarami & Chan, 2014; Iriana & Buttle, 2007).

Analytical CRM provides analysis of customer data and facilitates the understanding of individual behaviors and needs. Moreover, it enables customer behavior predictive modeling and purchase pattern recognition. Analytical tools such as data mining, data warehouses, and analytical processing are used by analytical CRM (Khodakarami & Chan, 2014).

1.2.3 CRM in the Hospitality Industry

CRM is a client-centric process that focuses on the relationship with the guest (Wu & Chen, 2012). It contributes to increasing guest satisfaction to loyalty and retention, lowering the acquisition costs, and increasing profits (Sarmaniotis et al., 2013; Lo et al., 2010; Rahimi, 2017). As a result, the goal of CRM is to gain knowledge of the client (needs, preferences, and emotions) to build an efficient connection (commercial and experiential) in each interaction. This is becoming more feasible due to advancements and innovations in information technology (IT) applications (González-Serrano et al., 2020).

The importance of CRM has increased in recent years, although it has been implemented for over thirty years in the hospitality industry (Sarmaniotis et al., 2013). This is due to the increased competition (Sigala, 2005), client's increased power and information as a result of digital environments (Anshari et al., 2019), and technological development (González-Serrano & Talón-Ballestero, 2020). Consequently, a hotel's profitability is inextricably linked to its ability to satisfy clients effectively and efficiently. CRM plays a vital role because it includes organizational procedures and strategies for better understanding its clients and applying this knowledge to the production and marketing of hotel services (Sharma, 2020).

Hotel CRM and technology have been studied in several works (Rahimi, 2017). First, to review what the systems consisted of, and which technologies were being used. Then, with the advent of the internet, to investigate how hotels used multimedia channels such as e-mail to build relationships with their guests. Recently, guests' engagement with social media has been explored (Ramos et al., 2017).

In the hotel sector, digital technology has generated a significant volume of client data sources in various formats, whose analysis is complex due to their diversity. We are dealing with organized data in traditional databases (from property management and CRM systems), as well as semi-structured and unstructured data (obtained from meta-search generated data, such as those given by Kayak, Trivago, and TripAdvisor or extracted from social networks, such as Facebook, Twitter or LinkedIn) (Talón-Ballestero et al., 2018; González-Serrano et al., 2020). Big Data technologies and approaches make it easier to access, store and analyze large volumes of client data gathered and processed at fast speeds (Anshari et al., 2019).

1.3. Big Data

The Big Data (BD) concept first appeared in the late 90s in computer science literature (Mariani, 2019). However, its first definition was only given by Laney (2001), identifying the three V's – Volume (data size in the order of Zettabytes), Velocity (rapidity of data generation, alteration, and transfer), and Variety (data can take on many formats/structures). Later, the definitional model was refined by including the Vs of Value (the process of extracting valuable knowledge from data via BD analytics) (Gantz & Reinsel, 2011) and Veracity (data governance concerning reliability), resulting in the creation of a 5Vs framework (Bello-Orgaz et al., 2016). Big Data consists of large volumes of data that are either structured or unstructured (Rivera, 2020) and conceptualized as "… high-volume, and high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation." (Gartner, 2018).

According to Li et al. (2018), data sources can divide big data into three categories: User-generated content (UGC) data (online textual and photo data actively submitted by users); device data (collected passively by devices such as GPS data, mobile roaming data, Bluetooth data, RFID data, WIFI data, and so forth) and transaction data (web search data, webpage visiting data, online booking data, and so on).

The hospitality industry has evolved into an information-intensive sector, with massive data volumes kept and practical applications that are not widely used. However, with the advent of Big Data, it is now possible to handle such data to achieve objectives and transform information into knowledge (Xiang et al., 2015; Talón Ballestero et al., 2018).

1.3.1 Big Data Analytics

In recent years, companies have increasingly relied on innovative software solutions to manage workloads, sustain profitability, and ensure competitiveness within their respective industries (Liebowitz, 2016). Business analytics is the science and practice of using quantitative data to decision-making and evaluating historical data to forecast business trends (Miles et al., 2013). It employs statistical analysis, data mining, text mining, and quantitative analysis (Azam & Tanweer, 2017). Data mining comprises methods that go beyond counts, descriptive techniques, and rule-based business procedures. It involves statistical and machine-learning methods that aid in decision-making (Shmueli et al., 2016). It is the process of extracting and analyzing massive volumes of data from different concepts and scenarios to discover hidden patterns in the provided dataset (Maheshwari, 2014). Text mining is the process of extracting interesting and non-trivial patterns from unstructured text sources to uncover knowledge from textual databases (Tan, 1999). Text mining is fundamentally unstructured compared to data mining, including information retrieval, text analysis, extraction, clustering, categorization, visualization, database technology, and machine learning (Aggarwal & Zhai, 2012; Lee et al., 2020).

Big data analytics eases the segmentation and more profound knowledge of the clients (Talón Ballestero et al., 2018). Continuous profile updating allows the hotel to interact with the client in real-time while also providing knowledge of the client's value and life cycle (Durson & Caber, 2016). Moreover, services can be modified to customer demands, enabling the delivery of individualized services and products. In addition, it allows carrying out marketing campaigns targeted at each client segment while predicting client reaction to those initiatives (González-Serrano et al., 2019).

2. Methodology

CRISP-DM (cross-industry process for data mining) methodology will be used in this work. It is a robust and well-proven methodology that provides a structured approach to planning a data mining project. Moreover, it is a popular business-oriented methodology for increasing the success of data mining projects. Its life cycle encompasses six phases - business understanding, data understanding, data preparation, modeling, evaluation, and deployment (Figure 1). The initial phase focuses on the problem from a business perspective and involves understanding the objectives and setting a data mining goal. The second phase regards the initial data collection and early data evaluation, including identifying data quality problems and finding patterns. The third phase prepares the data for modeling and consists of selecting, cleaning, integrating, and formatting the data to construct the final dataset. The fourth phase is when modeling techniques are applied to the data to create models. The already built model is reviewed and evaluated in the fifth phase to meet the business objectives. Finally, the last or deployment phase, presents the knowledge gained in a manner that the customer can use it. (Smart Vision Europe, 2021).

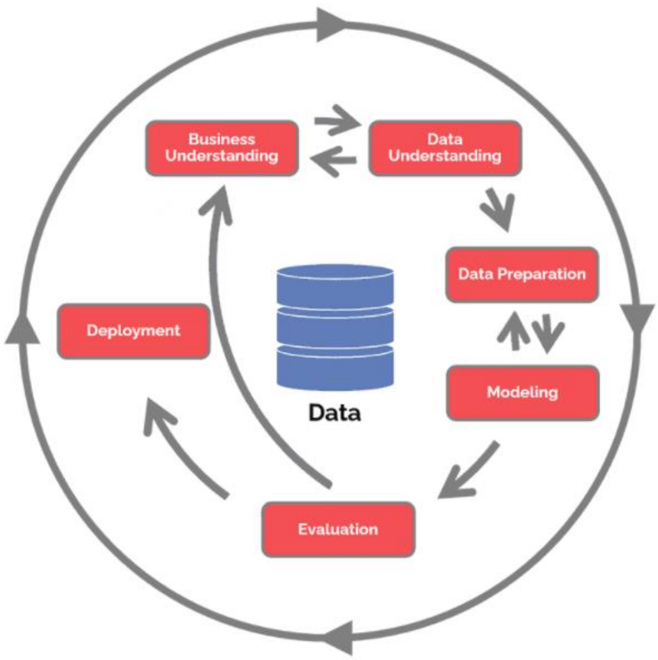

Figure 1 - CRISP-DM life cycle
Source: Data Science Process Alliance

2.1. Business Understanding

Hotel Group X's loyalty program is part of the group's CRM global strategy, and one of its main objectives is to increase repeat purchases. However, it seems that the outcomes have been disappointing. In fact, during the last three years, from the 1.4 million members, only 14.3% made one reservation, and only 3.6% have booked a room more than once. The company conducted an overall review of the program, and some weaknesses and flaws were pointed out:

1. The conversion rate (how much each point is worth in euros) is low, making it difficult for the clients to redeem their points. Guests need to consume a lot to gain enough points to exchange for rewards.
2. There is a scale drawback. Compared to larger hotel groups like Marriott or InterContinental, where there are thousands of alternatives for the guests to choose from, Hotel Group X has a few dozen hotel units, making the alternatives relatively limited. For instance, when people travel somewhere, the chance of coming across an InterContinental is substantially higher than a Hotel Group X's hotel unit. Ultimately, large hotel group's guests may choose the destination based on the existence of a hotel.
3. The benefits offered to the members are not clear. Sometimes even the staff is not fully aware of them, so the guests end up not experiencing them at all.
4. The service lacks consistency from one hotel to another regarding special treatment. For instance, the same guest visiting two different hotels from the group may be offered free water bottles during his stay in the first hotel but have no offers in the second one.
5. There are strictly limited partnerships and options to redeem points or to get discounts from.
6. Strategy-wise, the company focused on growing a customer base and promoting direct sales. Guests were given 10% and 15% off in reservations if they joined the loyalty program. This strategic move may have resulted in one-time reservations and no loyalty at all.

All these problems and decisions led to what is believed to be the bigger picture - guests do not recognize the program's value - and therefore, join the program to enjoy the direct discounts but, in the long run, end up not committing to it. As a result, the percentage of active customers who truly engage in the program's functions, such as making reservations, point accumulation, and reward redemptions, is low.

One of the company's main challenges is to revitalize the loyalty program. Nevertheless, despite all the group's efforts to restructure it and increase its effectiveness, no successful attempt has been made yet.

All the issues mentioned above are critical for the future success of the loyalty program. However, given the difficulty to measure and quantify some of these problems and project time constraints, it would not be possible to address all of them individually. As such, based on the data available, it was decided that this project would study the low redemption rate that is a consequence of the weaknesses and flaws of the program.

Therefore, by identifying active loyal customers (i.e., customers who engage in reward redemptions), this project aims to provide the company with insights about who these guests are now and who could eventually turn into one in the future, so that the group can design marketing strategies accordingly. The primary data mining objectives of this project are:

1. Uncover the most important predictors of reward redemptions.
2. Find out patterns and common characteristics among customers who engage in reward redemptions.

Furthermore, this project will also analyze the characteristics of members of the loyalty program who have made at least one reservation in the last three years but have not engaged in reward redemptions. This additional analysis aims to define a profile of these guests and compare them with the ones who redeemed rewards.

2.2. Data Understanding

The data of this project was extracted from an SQL server that contains customer data previously obtained by the hotel management software - Opera. There were extracted two different datasets in .csv format.

The first one (Dataset 1) consists of room reservations made by loyalty program members for any hotel unit, between the 1st of January 2018 and the 31st of December 2019. There were extracted 329193 records containing information about the reservation itself and customer-related information such as demographics and buying behavior. There was total number of 30 variables (Table 1).

Variable ID	Type	Measure	Description
Customer_FullName	String	Nominal	Guest's name
Customer_Email	String	Nominal	Guest's email
Customer_Country	String	Nominal	Guest's nationality
ID	Numeric	Scale	Guest's identification number
Customer_CRMCode	Numeric	Nominal	Guest's identification number (CRM system)
Profile_OpeCode	Numeric	Nominal	Guest's identification number (Opera system)
Customer_CodCard	Numeric	Nominal	Loyalty card identification number
Customer_BirthDate	Date	Scale	Guest's birth date
Reservation_ResortOpe	String	Nominal	Hotel booked
Reservation_ResortOpe	String	Nominal	Status of the reservation (checked-out or canceled)
Reservation_InternalIdentification	Numeric	Nominal	Reservation identification number
Reservation_Identification	Numeric	Nominal	Reservation identification number
Reservation_BookingDate	Date	Scale	Reservation date
Reservation_CheckInDate	Date	Scale	Check-in date
Reservation_CheckOutDate	Date	Scale	Check-out date
Reservation_BookingWindowDays	Numeric	Scale	Time between the reservation and the check-in (in days)
Reservation_Market	String	Nominal	Guest's segment
Reservation_Source	String	Nominal	Guest's segment

Reservation_Origin	String	Nominal	Guest's segment
Reservation_GroupMarketSegment	String	Nominal	Guest's segment
Reservation_MarketSegment	String	Nominal	Guest's segment
Card_Tier	String	Nominal	Guest's card tier
Card_CustomerCreatedOn	Date	Scale	Date of entry in the loyalty program
Reservation_AmtRoomRevenue	Numeric	Scale	Price paid for the room
Reservation_AmtFoodRevenue	Numeric	Scale	Amount spent in F&B
Reservation_AmtOtherRevenue	Numeric	Scale	Other expenditures (e.g., spa)
Reservation_QtyAdults	Numeric	Scale	Number of adults in the reservation
Reservation_QtyChildren	Numeric	Scale	Number of children in the reservation
WeekDays	Numeric	Scale	Number of weekdays (from Sunday to Thursday)
WeekendDays	Numeric	Scale	Number of weekend days (Friday and Saturday)

Table 1 - Dataset 1

Hotel Group X uses Opera, a property management system, to automate services such as the front office. Each reservation is then introduced into the system and a profile is created for the guest, if it had not been before. In addition, opera generates identification numbers for each guest and each room (i.e., reservation). That is why Table 1 contains so many kinds of identification variables. Moreover, if the guest joins the loyalty program, one more identification number is generated and introduced in the Opera guest profile.

Hotel Group X's customer segmentation process is quite complex. Guests are segmented into six market segments - Transient Direct, Transient Contracted, Transient Corporate, Group Business, Complimentary, and House Use. However, these belong to a broader segmentation. Other steps precede this ultimate segmentation. These include more detailed information about the booking channel (whether it is online or offline, direct or indirect), the reason of the stay, who paid for the trip, whether it is a promotional campaign, among many others. Every piece of information is given a code regarding “market”, “source”, and “origin”. There are multiple combinations of codes that result in another segmentation that precedes the broader one mentioned above. Table 2 provides the description and examples of each of the two final segmentation stages.

"Reservation_GroupMarketSegment"	**"Reservation_MarketSegment"**	**Description**
Transient Direct	TD - Site	Reservations via Hotel Group X's website
	TD - Central Reservations Direct Contact Center	Reservations via contact center
	TD - Local Reservations & Sales Office (Hotel or Area)	Reservations via the sales office
	TD - Affiliated Programs	Reservations via affiliated partners
Transient Contracted	TO - Negotiated	Reservations via offline tour operators
	TO - Standard	Reservations via offline tour operators
	TO - Web-Based B2B	Reservations via online B2B tour operators
	TO - Web-Based B2C	Reservations via online B2C tour operators
	TO - Specialist (General)	Reservations via specialized tour operators (e.g., Golf)
Corporate	TC - Negotiated	Business travelers with partnerships
	TC - Corp.	Outside employees on duty
	TC - Standard	Business travelers with no partnerships
	TC - Overbooking	When other hotels from the group are fully booked
Group Business	GB - Airline	Groups of airline crew members
	GB - Affiliated	Groups from affiliated partners
	GB - Congress	Groups from events outside the hotel (e.g., web summit)
	GB - Specialist Interest Groups	Groups from events inside the hotel (e.g., weddings)
	GB - Layover/Night Stops	Groups from exceptional situations (e.g., canceled flights)
	GB - Leisure	Groups with leisure purposes
	GB - MICE	Groups from events inside the hotel (e.g., congress)
	GB - Corp.	Groups from the company's meetings
	GB - Tour Series	Groups with leisure purposes from travel agencies
Complimentary	CP - Administration	Reservations for board members
	CP - DGO/GM	Reservations for employees

	CP - Fam Trips/Educational's	Reservations for partners
	CP - Residence / Owner	Property owners
	CP - Sales & Marketing	Offered by the marketing and sales departments (e.g., journalists)
House Use	House Use	Reservations for employees in duty

Table 2 - Segments' descriptions

The second dataset (Dataset 2) regards the loyalty card movements. There are three types of card movements – consumption, redemption, and expiration. Consumption happens when the client spends money in the hotel and earns points. It can be a room reservation, a meal, or any other expenditure. It is automated, the client does not need to ask for the points to be debited on his card. Redemption is a voluntary action where the client intends to exchange accumulated points for a reward. Finally, expiration is when the client does not redeem any reward for an extended period, so the points expire.

There are only two variables in this dataset – "Customer_CardCode" and "Type_Event" – where the former indicates the customer's ID, and the latter specifies the card's movement.

Dataset 2 includes all the card movements made by guests during the same period as Dataset 1. Thus, a total of 1618418 events were extracted (Table 3).

Variable ID	**Type**	**Measure**	**Description**
Customer_CodCard	Numeric	Scale	Loyalty card identification number
Type_Event	String	Nominal	Card movement (consumption, redemption, or expiration)

Table 3 - Dataset 2

2.3. Data Preparation

For this phase, an SPSS Statistics file and an Excel file were created to clean, select, and derive new attributes from the most relevant data. SPSS Modeler was later used for the integration step.

2.3.1. Variables Excluded

Due to privacy concerns, the names and emails of the guests were removed from the final dataset. From all the identification variables - whose sole function was to pair each reservation with a specific guest – only “Customer CodCard” was not excluded. Both datasets share this variable, so it will be required for the merge. All the others - “ID”, “CustomerCRMCode”, “Profile_OpeCode”, “Cutomer_CodCard”, “Reservation_InternalIdentification”, and “Reservation_Identification” - were excluded due to their uselessness for the project.

Moreover, four other variables were excluded: “Card_CustomerCreatedOn” due to its uselessness for this project; “Reservation_BookingDate” because there is other variable for the same purpose; “Customer_BirthDate” due to excessive missing values and “Reservation_Status” because the initial dataset did not contain any canceled reservation.

Finally, as stated above, the variables “Reservation_MarketSegment” and “Reservation_GroupMarketSegment” are used by Hotel Group X to segment the customers based on three other variables – “Reservation_Market”, “Reservation_Source”, “Reservation_Origin”. Therefore, it would be redundant to keep all these variables. The variable “Reservation_GroupMarketSegment” is quite general and lacks differentiation between two very distinct channels – online and offline. Consequently, only the “Reservation_MarketSegment” was considered and will be used to create a new one.

2.3.2. Variables Created

Regarding Dataset 1, some new variables were created to aggregate the records in categories and facilitate the analysis.

Since the hotel group owns many distinct hotels located throughout the world, two new variables - “Hotel_Area” and “Hotel_Type” – were created from the variable “Reservation_ResortOpe” (name of the hotel).

The first divides hotels into four categories based on their location: Europe, Africa, South America, and North America. However, differentiation was crucial because the group owns many of the units and some of the most important markets in Portugal. Portugal has four key markets, each with its size, customer demographics, and seasonality. As a result, Lisbon, Porto, Madeira, and Algarve are also included in this classification. The remainder of Portugal (units not located in the regions mentioned above) has its own category. One hundred twenty-five different hotel units were reduced to nine new categories.

The second variable - “Hotel_Type” - categorizes the hotels into three new categories - Resort, City, and Lodge. This classification helps to differentiate the hotel units based on two distinctive factors – product and destination. For example, while resorts are associated with leisure and family holidays located in sunny places, city hotels are more directed to business travelers and shorter stays. On the other hand, lodges are smaller hotels located near historical centers. This classification is essential because the type of hotel may be related to the guests’ characteristics and behaviors and eventually be a predictor of reward redemptions.

It was decided to replace the variable "Reservation_MarketSegment" with "Market_Segment" (the former was then eliminated). This segmentation was addressed with the company. It is felt that the following is the most accurate way to segment customers depending on their markets and booking channels - Site, Contact Center, Other Direct, Contracted Online, Contracted Offline, Corporate, Groups, and Other. This segmentation considers the importance of the group's two primary direct channels and the requirement to differentiate between online and offline channels. On the other hand, “House Use” and “Complimentary” were not relevant enough to be categories in this project. Moreover, there was a need to shorten the number of classes in the variable to facilitate the analysis. Once more, this categorization is crucial for the study since it is believed that the customer segment could also predict reward redemptions.

Finally, the variable “Lenght_Of_Stay”, which refers to the number of days that a guest stays in the hotel, was computed from the variables “Reservation_CheckInDate” and “Reservation_CheckOutDate”. “Lenght_Of_Stay” is also believed to have an impact on customers’ loyalty intentions.

Table 4 shows all the created and eliminated variables as well as the ones that were kept and whose names have been changed.

Old Variables	Situation	New Name
Customer_FullName	Eliminated	N/A
Customer_Email	Eliminated	N/A
Customer_Country	Kept	Customer_Country
ID	Eliminated	N/A
Customer_CRMCode	Eliminated	N/A
Profile_OpeCode	Eliminated	N/A
Customer_CodCard	Kept	Card_No
Customer_BirthDate	Eliminated	N/A
Reservation_ResortOpe	Eliminated	N/A
Reservation_Status	Eliminated	N/A
Reservation_InternalIdentification	Eliminated	N/A
Reservation_Identification	Eliminated	N/A
Reservation_BookingDate	Eliminated	N/A
Reservation_CheckInDate	Eliminated	N/A
Reservation_CheckOutDate	Eliminated	N/A
Reservation_BookingWindowDays	Kept	Booking_Window
Reservation_Market	Eliminated	N/A
Reservation_Source	Eliminated	N/A
Reservation_Origin	Eliminated	N/A
Reservation_GroupMarketSegment	Eliminated	N/A
Reservation_MarketSegment	Eliminated	N/A
Card_Tier	Kept	Card_Tier
Card_CustomerCreatedOn	Eliminated	N/A
Reservation_AmtRoomRevenue	Kept	Room_Rev
Reservation_AmtFoodRevenue	Kept	Food_Rev
Reservation_AmtOtherRevenue	Kept	Other_Rev
Reservation_QtyAdults	Kept	Adults_Qty
Reservation_QtyChildren	Kept	Children_Qty
WeekDays	Kept	Week_Days
WeekendDays	Kept	Weekend_Days
New Variables	**Situation**	**New Name**
Lenght_Of_Stay	Created	N/A
Hotel_Area	Created	N/A
Hotel_Type	Created	N/A
Market_Segment	Created	N/A

Table 4 - Eliminated, created and edited variables

2.3.3. Dummy Variables

Dataset 1 contains data about every reservation made by the loyalty program members between 2018 and 2019. However, each line in the dataset represents a single reservation rather than a single guest. In accordance with the project's goal, all reservations made by a guest had to be aggregated so that the guest’s attributes could be analyzed. Therefore, dummy variables were created for the nominal variables likely to change from one reservation to another. For instance, if a guest books a room in Lisbon and, in his next stay, books a room in Algarve, these two reservations will have different attributes in at least one variable (“Hotel_Area”). Dummy variables will incorporate all the reservation attributes for every reservation made by one guest. In the example above, if the guest first stays in Lisbon, the new variable “Area_Lisbon” will get the score “1” while all the other dummy variables of the hotel areas will be left with a “0”. The same happens in the second stay, this time with “Area_Algarve”. There were created dummy variables for the variables “Hotel_Area”, “Hotel_Type” and “Market_Segment”. All the other ones either relate to customer’s characteristics that will not change regardless of the reservation or contain numerical values that can be computed.

2.3.4. Excluded Records

A data audit was performed on the initial dataset to ensure its quality. It was found that it contained several odd values, most likely because of front-office mistakes. These were significant amounts for a single reservation worth thousands of euros as well as negative sums for lodging or meal revenue. To obtain the most accurate results, all the negative values were removed from the following variables: "Length_Of_Stay," "Booking_Window," "Week_Days," "Weekend_Days," "Room_Rev," "Food_Rev," and "Other_Rev." Furthermore, a lower boundary of 50€ and maximum boundaries of 10000€, 1000€, and 200€ were established for “Room_Rev,” “Food_Rev,” and “Other_Rev,” respectively. These values were discussed with the company, based on average revenues and guided by common sense and reasonableness criteria.

2.3.5. Aggregation

As explained above, a readjustment of the dataset had to be done. SPSS Modeler aggregate node was used to aggregate all the reservations of a member in a single line.

In Dataset 1, records were aggregated by the variable “Card_No” (key field), the customer identification. Hence, another variable was created and labeled “Card_No_Count”, indicating the number of times a guest stayed in the hotel units. Records regarding the other variables were aggregated differently. The mean was computed for each ordinal or scale variable. As for the nominal ones, the min (minimum) was computed, meaning that they were left with the same value.

As for Dataset 2, the card track records were also aggregated by “Card_No” to compute redemption frequency per guest. Moreover, the select node only included the records containing “Redemption” in the variable “Type_Event”. This selection aligns with the project's purpose, which is to predict what attributes will influence the reward redemptions. Consumption, as stated before, is an automatic and involuntary process, therefore it does not portray engagement or loyalty intentions towards the brand. Unfortunately, only 8049 records were selected out of 1357680 card activity track records (including consumption, expiration, and redemption), suggesting that the redemption rate is very low. A new variable called “Loyalty_Score” was generated containing the number of times each guest redeemed rewards.

2.3.6. Merge

Both datasets were then ready to be merged one to another. The merge node from SPSS Modeler was used for this task. The key for the merge was the “Card_No” variable. Matching and non-matching records (full outer join) were included in this merge. Some of the records from Dataset 1 were not present in Dataset 2. Those were the guests that did stay at one of the group’s hotels but ended up not using the card for any redemption. Therefore, in the merged file, those records got the value “null” and were replaced by a value of "0". On the other hand, some records were present in Dataset 2 but not in Dataset 1. These records accounted for previously excluded records due to odd values (e.g., negative revenues) and were discarded from the merged file.

2.3.7. Target Variable and Data Balance

Two final adjustments had to be done before running the model. The data mining goal of this project is to predict the variables that influence reward redemptions the most. Therefore, the target variable will be "If_Loyalty", describing whether the customer ever redeemed a reward. Through the derive node, guests who never redeemed rewards got the value "0" and those who had ever redeemed rewards, regardless of the number of times, got the value "1". The variable "Loyalty_Score" (the number of times each guest redeemed rewards) previously created was excluded after conducting the descriptive statistics analysis.

Moreover, due to the limited number of reward redemption records, the dataset had to be balanced. Data should have approximately equal numbers of both outcomes (0 or 1), so the model has a better chance of finding patterns that distinguish the two groups (Chawla et al., 2004). The balance node was used for this task, and the condition "If_Loyalty=0" was given the factor of 0.05. There were selected 14213 records, where 66.36% were guests that did not redeem rewards ("If_Loyalty"=0) and 33.64% were guests who redeemed ("If_Loyalty"=1).

The final dataset contains 34 variables as shown in Table 5. The total number of guests is 14213.

Variable ID	**Type**	**Measure**	**Role**	**Description**
Card_No	Numeric	Nominal	Input	Guest's ID
Card_No_Count	Numeric	Scale	Input	Number of stays
Coustomer_Country	String	Nominal	Input	Guest's nationality
Lenght_Of_Stay_Mean	Numeric	Scale	Input	Average time per stay (in days)
Booking_Window_Mean	Numeric	Scale	Input	Average time in advance the reservation was made (in days)
Week_Days_Mean	Numeric	Scale	Input	Average weekdays per stay (in days)
Weekend_Days_Mean	Numeric	Scale	Input	Average weekend days per stay (in days)
Card_Tier_Min	String	Nominal	Input	Guest's card tier
Adults_Qty_Mean	Numeric	Scale	Input	Average adults per stay

Children_Qty_Mean	Numeric	Scale	Input	Average children per stay
Room_Rev_Mean	Numeric	Scale	Input	Average amount paid per reservation per stay (in euros)
Food_Rev_Mean	Numeric	Scale	Input	Average amount spent in F&B per stay (in euros)
Other_Rev_Mean	Numeric	Scale	Input	Average amount spent in other expenditures per stay (in euros)
Area_Portugal_Mean	Numeric	Scale	Input	Average stays in Portugal (1-all of them; 0-none)
Area_Lisbon_Mean	Numeric	Scale	Input	Average stays in Lisbon (1-all of them; 0-none)
Area_Porto_Mean	Numeric	Scale	Input	Average stays in Porto (1-all of them; 0-none)
Area_Madeira_Mean	Numeric	Scale	Input	Average stays in Madeira (1-all of them; 0-none)
Area_Algarve_Mean	Numeric	Scale	Input	Average stays in Algarve (1-all of them; 0-none)
Area_Europe_Mean	Numeric	Scale	Input	Average stays in Europe (1-all of them; 0-none)
Area_Africa_Mean	Numeric	Scale	Input	Average stays in Africa (1-all of them; 0-none)
Area_South_America_Mean	Numeric	Scale	Input	Average stays in South America (1-all of them; 0-none)

Area_North_America_Mean	Numeric	Scale	Input	Average stays in North America (1-all of them; 0-none)
Type_Resort_Mean	Numeric	Scale	Input	Average stays in a Resort Hotel (1-all of them; 0-none)
Type_City_Mean	Numeric	Scale	Input	Average stays in a City Hotel (1-all of them; 0-none)
Type_Lodge_Mean	Numeric	Scale	Input	Average stays in Lodge Hotel (1-all of them; 0-none)
Market_Direct_Site_Mean	Numeric	Scale	Input	Average reservations via the company's website (1-all of them; 0-none)
Market_Direct_Other_Mean	Numeric	Scale	Input	Average reservations via other direct channels (1-all of them; 0-none)
Market_Direct_Contact_Center_Mean	Numeric	Scale	Input	Average reservations via contact center (1-all of them; 0-none)
Market_Contracted_Online_Mean	Numeric	Scale	Input	Average reservations via contracted online (1-all of them; 0-none)
Market_Contracted_Offline_Mean	Numeric	Scale	Input	Average reservations via contracted offline (1-all of them; 0-none)
Market_Corporate_Mean	Numeric	Scale	Input	Average reservations via corporate agreements (1-all of them; 0-none)

Market_Groups_Mean	Numeric	Scale	Input	Average reservations via group agreements (1-all of them; 0-none)
Market_Other_Mean	Numeric	Scale	Input	Average reservations via other ways (1-all of them; 0-none)
If_Loyalty	Numeric	Nominal	Target	Reward redemptions (1-yes; 0-no)

Table 5 - Final dataset

2.3.8. Descriptive Statistics

Descriptive statistics will be used to find patterns and common characteristics among the loyalty program members. These techniques provide basic information about the sample. The conclusions withdrawn in this phase will only be the initial part of extensive analysis in the modeling phase (Hand et al., 2001).

IBM SPSS Modeler select node was used twice to distinguish the customers who ever redeemed a reward from those who never did. The statistics node was used to compute the mean, min, max, and standard deviation of 30 variables. Moreover, the distribution node was used to display frequency graphs of two other variables.

Figure 2 shows the distribution of the variable “If_Loyalty” and indicates that only 2.49% of the guests redeemed rewards between January 2018 and December 2019. This percentage confirms the very low redemption rate acknowledged by the company.

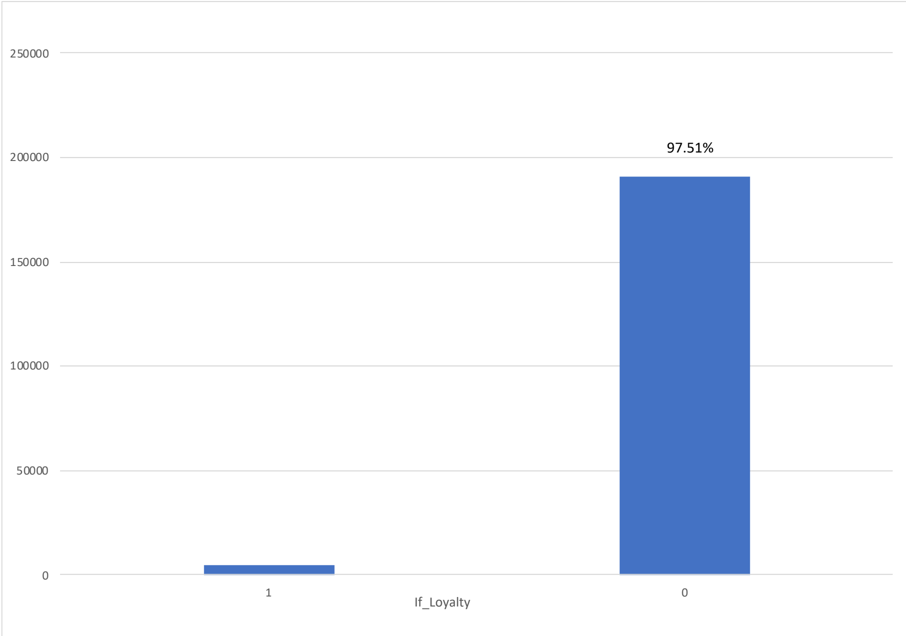

Figure 2 - Distribution of the variable “If_Loyalty”

Figure 3 shows the distribution of the variable “Loyalty_Score”. Most of the guests (66.41%) only redeemed rewards once, and the percentage of guests redeeming rewards more than five times was just 1.65%. Therefore, it was decided to group all the records greater than 5 in a single category (“<5”) through the derive node from SPSS Modeler.

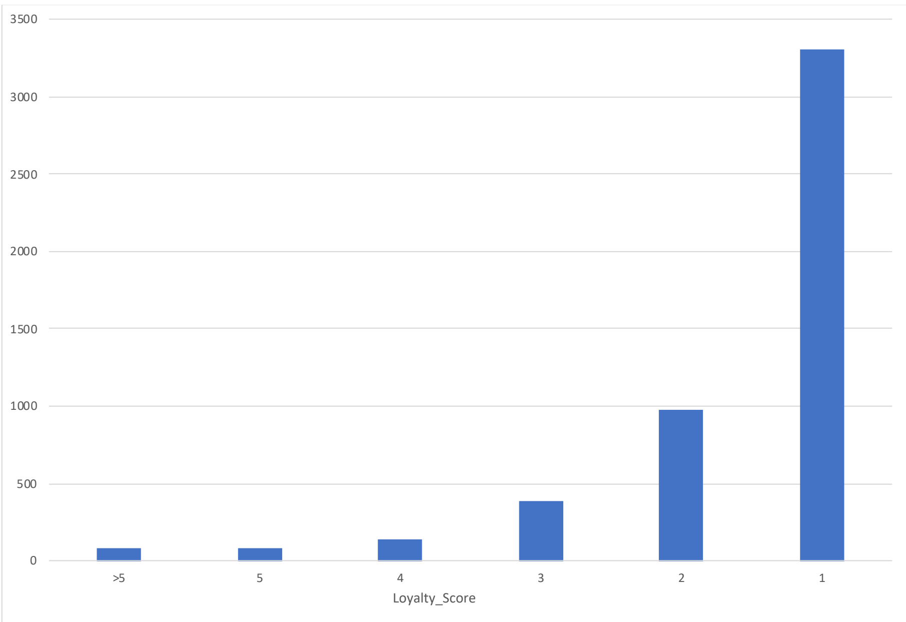

Figure 3 - Distribution of the variable “Loyalty_Score”

The following analysis is based on four levels: guest traveling behavior, geographic area where they travel to, market (mainly the booking channel), and the type of hotel they decided to stay in. Moreover, it contrasts the customers who have made a reward redemption at least once (If_Loyalty=1) with those who have never done it (If_Loyalty=0).

Table 6 suggests that guests who did not redeem rewards stayed, on average, 1.43 times at Hotel Group X. Their loyalty score is 0.000 since they never redeemed rewards. Moreover, they stayed an average of 3.98 days (2.74 weekdays and 1.23 weekend days). On average, they make their reservation 51.23 days in advance. As for the expenditures, they spend, on average, 446.41€, 112.78€ and 7.19€ in the room, F&B and others, respectively. Each reservation includes, on average, 1.929 adults and 0.22 children.

Conversely, guests who have redeemed rewards at least once, booked, on average, 5.52 times. They redeemed rewards, on average, 1.62 times. As for the duration of their stay, the average is 3.55 days (2.41 weekdays and 1.14 weekend days). They book on average 60.73 days in advance. As for the expenditures, they spend, on average, 482.40€, 118.93€ and 5.84€ in the room, F&B and others, respectively. Finally, each reservation includes, on average, 1.98 adults and 0.36 children.

Buying Behavior	**If_Loyalty=1**		**If_Loyalty=0**	
Variable	Mean	SD	Mean	SD
Card_No_Count	5.82	7.53	1.43	1.25
Loyalty_Score	1.62	1.20	0.00	0.00
Lenght_Of_Stay_Mean	3.55	2.66	3.98	3.17
Booking_Window_Mean	60.73	60.25	51.23	62.09
Week_Days_Mean	2.41	2.02	2.74	2.43
Weekend_Days_Mean	1.14	0.83	1.23	1.04
Adults_Qty_Mean	1.98	0.38	1.92	0.44
Children_Qty_Mean	0.36	0.60	0.22	0.55
Room_Rev_Mean	482.4	452.98	446.41	505.14
Food_Rev_Mean	118.9	112.73	112.78	142.03
Other_Rev_Mean	5.84	15.77	7.19	23.32

Table 6 - Descriptive statistics of the variables regarding Buying Behavior

Table 7 shows the preferences of the guests who did not redeem rewards regarding the geographic area of the hotel where they stayed. The analysis revealed that the area with the highest mean and, therefore, the most preferred one for the members, is Portugal (please recall that this variable only regards the hotel units that are not specifically located in Lisbon, Porto, Algarve, and Madeira), with a mean of 0.37. The min and max are 0 and 1, respectively, due to the dummy variables). This means that, on average, guests stay in Portugal, 37% of their total stays at Hotel Group X. Madeira and Algarve are also preferential areas for the guests, with means of 0.28 and 0.10, respectively.

As for guests who redeemed rewards, Table 7 shows that guests stay in Portugal, on average, 44% of their stays at Hotel Group X. On average, 27.3% of their holidays are in Algarve and 12.2% in Madeira. All the other destinations are significantly less important.

Hotel Area	**If_Loyalty=1**		**If_Loyalty=0**	
Variable	Mean	SD	Mean	SD
Area_Portugal_Mean	0.44	0.40	0.37	0.47
Area_Lisbon_Mean	0.02	0.10	0.03	0.17
Area_Porto_Mean	0.03	0.11	0.08	0.27
Area_Madeira_Mean	0.12	0.28	0.28	0.44
Area_Algarve_Mean	0.27	0.37	0.10	0.30
Area_Europe_Mean	0.07	0.23	0.07	0.25
Area_Africa_Mean	0.00	0.06	0.00	0.09
Area_South_America_Mean	0.02	0.12	0.02	0.13
Area_North_America_Mean	0.00	0.05	0.00	0.08

Table 7 - Descriptive statistics of the variables regarding Hotel Area

Table 8 presents guest’s preferences regarding hotel types. It suggests that resorts are the most preferred hotels for guests who did not redeem rewards, with a mean of 0.44, followed by lodges and city hotels, with 0.28 and 0.26, respectively.

Similarly, guests who redeemed rewards stay, on average, 44% of their visits in resorts. The means of lodge and city hotels are 0.36 and 0.18, respectively.

Guest_Market	**If_Loyalty=1**		**If_Loyalty=0**	
Variable	Mean	SD	Mean	SD
Market_Direct_Site_Mean	0.62	0.37	0.39	0.47
Market_Direct_Other_Mean	0.04	0.13	0.03	0.16

Market_Direct_Contact_Center_Mean	0.27	0.34	0.15	0.35
Market_Contracted_Online_Mean	0.01	0.08	0.18	0.38
Market_Contracted_Offline_Mean	0.00	0.05	0.16	0.36
Market_Corporate_Mean	0.03	0.15	0.02	0.14
Market_Groups_Mean	0.00	0.02	0.04	0.21
Market_Other_Mean	0.00	0.00	0.00	0.01

Table 8 - Descriptive statistics of the variables regarding Guest Market

Table 9 suggests that the market that stands out for guests who did not redeem rewards is the "Market_Direct_Site", with a mean of 0.39, representing those who make their reservations through the Hotel Group X's website. This means that, on average, guests book through the website 39% of the times. In addition, "Market_Direct_Contact_Center", "Market_Contracted_Online" and "Market_Contracted_Offline" are also big markets among these guests with means of 0.15, 0.18, and 0.16, respectively.

Similarly, "Market_Direct_Site" is the most significant market segment among customers who redeemed rewards, with a mean of 0.62, followed by "Market_Direct_Contact_Center", with 0.27. However, unlike guests who did not redeem rewards, the contracted markets do not stand out among guests who redeemed rewards. The third biggest market, but significantly less important, is the "Market_Direct_Other", with a mean of 0.04.

Hotel_Type	**If_Loyalty=1**		**If_Loyalty=0**	
Variable	Mean	STD	Mean	STD
Type_Resort_Mean	0.44	0.41	0.44	0.49
Type_City_Mean	0.18	0.32	0.26	0.43
Type_Lodge_Mean	0.36	0.39	0.28	0.44

Table 9 - Descriptive statistics of the variables regarding Hotel Type

2.4 Modeling

This phase focuses on selecting and building appropriate models for testing and analyzing the data. It includes four tasks - selecting the modeling techniques, designing tests, building the model, and assessing it (Chapman et al., 2000). Given the project's goal of predicting future reward redemptions, classification models should be used to determine which variables have the most impact on redemptions. Classification models attempt to draw conclusions from existing data and predict the value of one or more outcomes. C&RT was chosen in the first place. However, despite the good accuracy of the model, the results may have been insufficiently conclusive due to ambiguous splitting values. Consequently, a Logistic Regression model was added to the study. Both models were chosen due to their efficiency and simplicity even for non-data mining experts.

Data was split into two separate subsets for training (70%) and testing (30%) to get a good indication of how the model will generalize to larger, similar datasets (IBM, 2021). The input fields (predictors) were all the variables from the final dataset apart from “Card_No”, which is an identification variable, “Card_Tier_Mean” and “Card_No_Count”. The two last ones were excluded from the models due to their correlations between them and with redemptions. This redundancy would result in no information gain. A customer is expected to redeem more if he stays in the hotel more often because points are earned when guests make reservations. The greater the consumption, the greater the chance to redeem reward due to points accumulation. “Card_Tier_Min” follows the same logic. Upper-level tiers infers frequent consumption.

2.4.1 C&RT

SPSS Modeler C&R Tree node was used to apply the Classification and Regression Tree (C&RT) analysis to the data. C&RT “is a tree-based classification and prediction method” which “uses recursive partitioning to split the training records into segments with similar output field values”. It “starts by examining the input fields to find the best split, measured by the reduction in an impurity index that results from the split. The split defines two subgroups, each of which is subsequently split into two more subgroups, and so on until one of the stopping criteria is triggered. All splits are binary (only two subgroups)” (IBM, n.d, p. 83).

The SPSS Modeler default options were used in this model. The impurity measure was the Gini Index. The maximum tree depth was set to 5. Sometimes the tree can grow so large, almost to the point where it exactly fits the training data, with only one observation in each leaf. This, however, leads to overfitting and poor predictions on independent test sets. To determine an appropriate tree size, it creates an excessively huge tree until a specific minimum node size is reached. The tree should then be pruned back to its ideal size (Moisen, 2008). Therefore, the tree was pruned to avoid overfitting. As for the stopping rules, 2% and 1% were used for the minimum records in the parent and child branches, respectively.

2.4.2. Logistic Regression

Logistic Regression is another widely used, well-understood, and often well-performing supervised learning technique (Caie et al., 2021). A multiple Logistic Regression model has a dependent (outcome or response) variable that has two possible values (often coded with the values 0 and 1) and more than one independent variable (predictor variables). The purpose of the logistic analysis includes determining which predictors are important and how they affect the outcome, as well as creating a parsimonious and effective prediction equation (Elliott & Woodward, 2014).

In multiple Linear Regression the expected value of a response variable, y, is modeled as a linear function of the explanatory variables:

$$E(y) = \beta_0 + \beta_1 x_1 + \cdots + \beta_q x_q$$

The expected value for a binary response with the values 0 and 1 (failure or success) is simply the probability, p, that the variable takes the value 1, i.e., the probability of success. A more appropriate method is to model p indirectly using the logit transformation of p, i.e., ln[p/(1 – p)] (Landau & Everitt, 2004). This results in the Logistic Regression model:

$$ln \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \cdots + \beta_q x_q$$

To put it another way, the log-odds of success are represented as a linear function of the explanatory variables. In a Logistic Regression model, the estimated regression coefficients indicate the estimated change in log-odds corresponding to a unit change in the corresponding explanatory variable if the other explanatory variables remain constant. Typically, the parameters are exponentiated to get odds-based outcomes (Landau & Everitt, 2004). The Logistic Regression model may be expressed in terms of p as:

$$p = \frac{\exp\,(\beta_0 + \beta_1 x_1 + \cdots + \beta_q x_q)}{1 + \exp\,(\beta_0 + \beta_1 x_1 + \cdots + \beta_q x_q)}$$

The model was generated mainly using the SPSS default settings. The stepwise method of field selection was used. Stepwise builds the method in steps. Terms that have not yet been included in the model are examined at each stage, and if the best of those terms considerably improves the prediction power of the model, it is added. Simultaneously, existing terms in the model are reevaluated to see if any of them can be eliminated without significantly detracting from the model. The final model is created when no more terms can be added to improve the model and no more terms can be deleted without detracting from the model (IBM, n.d, p. 163).

2.5. Evaluation

2.5.1. C&RT

The model correctly predicted 83.26% of the cases in the training set with an AUC (area under the curve) value of 0.856 and a Gini Index of 0.701. As for the testing set, the model correctly predicted 83.41% of the cases with an AUC value of 0.852 and a Gini Index of 0.704. The fact that the correctly predicted values are noticeably high and very similar to each other is a good indicator of the high accuracy of the model. In addition, there are other relevant evaluation measures to assess the performance of this model - accuracy, precision, true positive rate (or recall), and true negative rate. These measures can be calculated with four variables - true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) - which construct the coincidence matrix (Figure 4 and Figure 5).

		True Class	
		Positive	Negative
Predicted Class	Positive	TP = 5967	FP= 664
	Negative	FN= 1018	TN= 2399

Figure 4 - Coincidence Matrix – C&RT

Measure	Meaning in this project
TP- Number of instances predicted as 1, which were actually 1	Number of instances that predicted that customer redeemed a reward, and he did.
TN- Number of instances predicted as 0, which were actually 0	Number of instances that predicted that the customer didn't redeem a reward and he didn't.

FP - Number of instances predicted as 1, which were actually 0	Number of instances that predicted that the customer redeemed a reward, but he didn't.
FN - Number of instances predicted as 0, which were actually 1	Number of instances that predicted that the customer didn't redeem a reward, but he did.

Note: 1- redeemed a reward; 2 - did not redeem a reward

Table 5 - Coincidence Matrix components

Accuracy (83.26%) represents the percentage of cases that the model correctly predicted among all the cases (Powers & Martin, 2011). True positive rate (85.43%) is the percentage of positive cases correctly classified from all the positive ones. The true negative rate (78.84%) represents the percentage of negative cases that are correctly classified. Finally, precision (89.99%) means the percentage of actual positive cases among all the cases classified as positive. All in all, the model is believed to be accurate at predicting the guest’s reward redemptions.

2.5.2. Logistic Regression

The model correctly predicted 70.29% of the cases (accuracy) in the training set with an AUC value of 0.772 and a Gini Index of 0.54. As for the testing set, the model correctly predicted 70.33% of the cases with an AUC value of 0.77 and a Gini Index of 0.54. The Logistic Regression, using a stepwise method of variable selection, explained 32.5% of the variance in redemptions (Nagelkerke *R2*).

Other measures including true positive rate, true negative rate and precision were also calculated for this model, with values of 74%, 58.62%, 84.88%, respectively.

Overall, although the values are lower than the ones achieved with the C&RT, they are still acceptable.

		True Class	
		Positive	Negative
Predicted Class	Positive	TP = 5670	FP= 1010
	Negative	FN= 1992	TN= 1431

Figure 5 - Coincidence Matrix – Logistic Regression

2.6. Deployment/Results

2.6.1. C&RT

Results showed that the most important predictor of the reward redemptions and significantly more important than all the others is the "Market_Direct_Site_Mean", with importance of 0.46. The second most important predictor is the "Area_Algarve_Mean", followed by the "Type_City_mean". All the other variables are less significant predictors of reward redemptions (Figure 6).

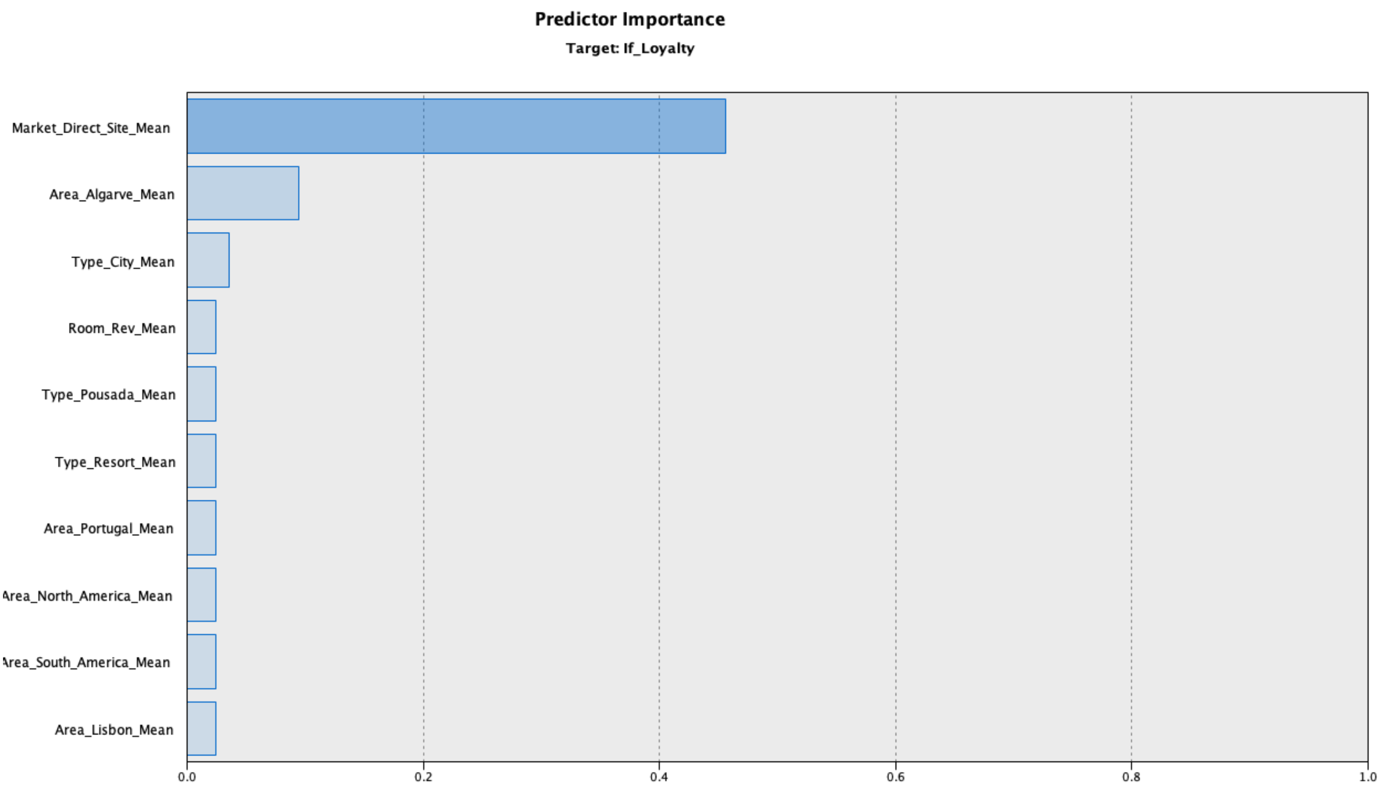

Figure 6 - Predictor importance

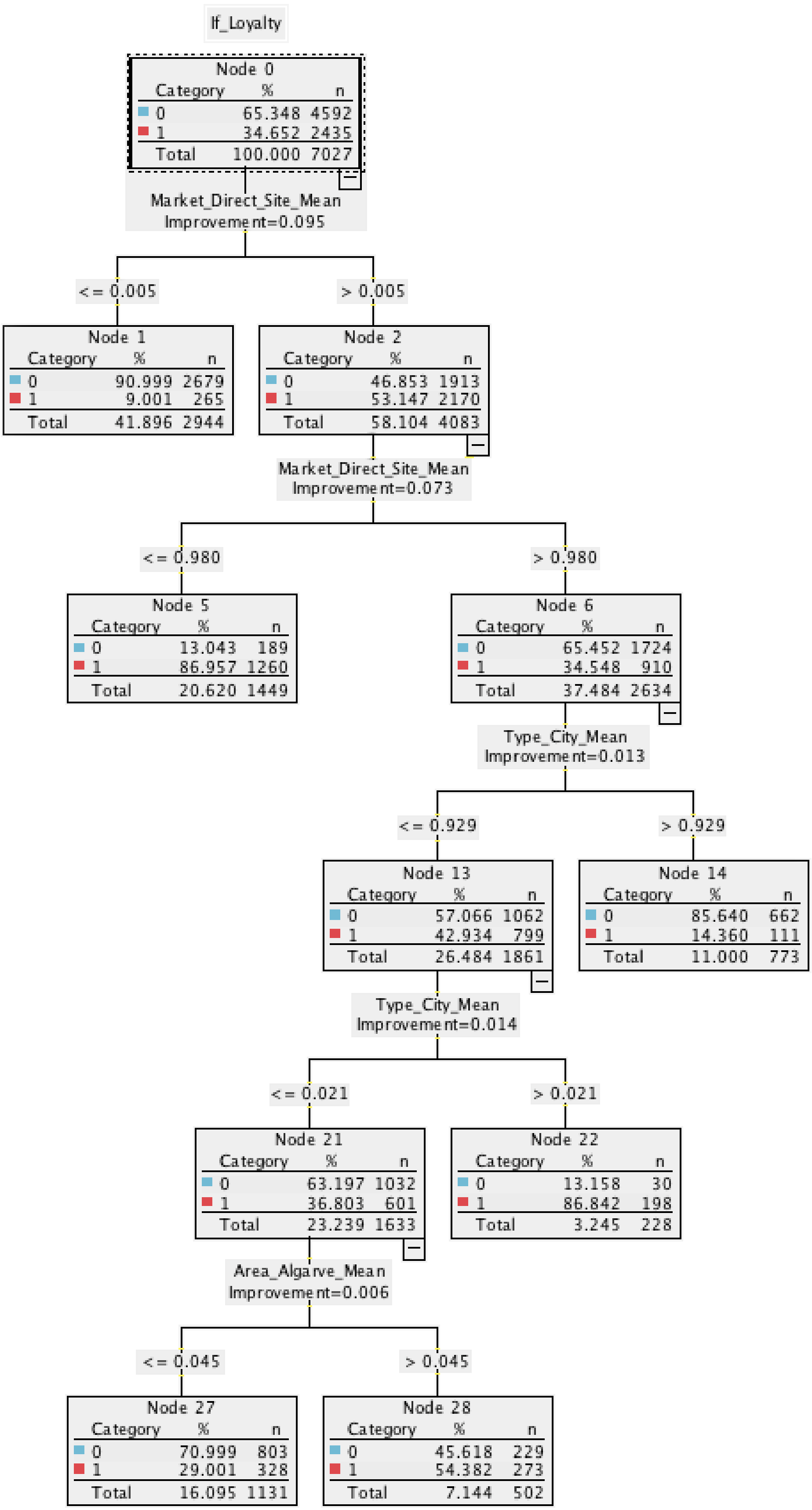

Figure 7 - C&RT - Decision Tree

The tree (Figure 7) contained 7027 observations and, as stated above, the most influential variable was “Market_Direct_Site_Mean” (which indicates that the booking channel used for the reservation was the Hotel Group X’s website).

The first and second splits of the tree are based on “Market_Direct_Site_Mean”. The tree suggests that 86.96% of the guests who booked via the company’s website (“Market_Direct_Site_Mean”) between 0.5% and 98% of their stays ended up redeeming rewards. Even though the interval of values in this node is extensive, guests in this group have definitely made at least one reservation through the website. Otherwise, the “Market_Direct_Site_Mean” value would have been 0. Recall that as a dummy variable, it ranges from 0 to 1, with 0 indicating that the guest made no reservations through the company's website and 1 implying that the guest made all reservations this way.

This tree's third and fourth splits depend on “Type_City_Mean”, which is also a dummy variable and therefore varies between 0 - from all the guest’s stays at Hotel Group X, none was in a city hotel - and 1 – implying that the guest stayed in city hotels all his visits. Results showed that 86.84% of the guests who made more than 98% of their reservations through the company’s website (“Market_Direct_Site_Mean”) and stayed city hotels (“Type_City_Mean”) between 2.1% and 92.9% of their visits, ended up redeeming rewards. Once again, the range is extensive, but it is possible to conclude that these guests stayed in a city hotel at least once.

The last split is based on “Area_Algarve_Mean”. The tree reveals that 54.38% of the guests who booked via the group’s website (“Market_Direct_Site_Mean”) more than 98% of their stays, who stayed in city hotels (“Type_City_Mean”) less than 2.1% of their visits and stayed in Algarve (“Area_Algarve_Mean”), more than 4.5% of the times, redeemed rewards.

On the other hand, 90.99% of the guests who used the company’s website as booking channel (“Market_Direct_Site_Mean”) for less than 0.5% of their visits, do not redeem rewards. In addition, 85.64% of the guests who booked via Hotel Group X’s website (“Market_Direct_Site_Mean”) more than 98% of the times and stayed in city hotels (“Type_City_Mean”) more than 92.9% of their visits, do not redeem rewards. Moreover, 70.99% of guests who booked via the company’s website (“Market_Direct_Site_Mean”) more than 98% of their stays, who stayed in city hotels (“Type_City_Mean”) less than 2.1% of their visits and stayed in Algarve (“Area_Algarve_Mean”) less than 4.5% of their visits, did not redeem rewards.

2.6.2. Logistic Regression

	B	Wald	Sig.	Exp(B)
Booking_Window_Mean	0.004	62.277	0.000	1.004
Room_Rev_Mean	.000	21.292	.000	1.000
Food_Rev_Mean	.002	60.742	.000	1.002
Other_Rev_Mean	-.004	7.308	.007	.996
Area_Portugal_Mean	.591	50.586	.000	1.805
Area_Porto_Mean	**-.761**	25.657	.000	.467
Area_Madeira_Mean	.539	14.975	.000	1.715
Area_Algarve_Mean	1.310	90.220	.000	**3.706**
Area_South_America_Mean	.492	6.745	.009	1.636
Type_Resort_Mean	-.261	6.289	.012	.771
Market_Direct_Site_Mean	4.388	362.475	.000	**80.493**
Market_Direct_Other_Mean	4.241	256.195	.000	**69.464**
Market_Direct_Contact_Center_Mean	4.246	328.999	.000	**69.853**
Market_Contracted_Online_Mean	1.257	21.845	.000	**3.514**
Market_Corporate_Mean	4.596	295.262	.000	**99.099**

Table 6 - Logistic Regression - Parameter Estimation

The main output for the Logistic Regression model can be viewed in Table 11. The selected variables are – “Booking_Window_Mean”, “Room_Rev_Mean”, “Food_Rev_Mean”, “Other_Rev_Mean”, “Area_Portugal_Mean”, “Area_Porto_Mean”, “Area_Madeira_Mean”, “Area_Algarve_Mean”, “Area_South_America_Mean”, “Type_Resort_Mean”, “Market_Direct_Site_Mean”, “Market_Direct_Other_Mean”, “Market_Direct_Contact_Center_Mean”, “Market_Contracted_Online_Mean” and “Market_Corporate_Mean”. All these variables are statistically significant (sig.<0.05).

Logistic Regression presupposes that the increasing variables with positive coefficients result in an increasing likelihood of redeeming redemptions. On the other hand, increasing variables with a negative coefficient result in a decreasing likelihood of reward redemptions.

The results show that there is a significant predictive power for the variable “Matket_Corporate_Mean” which has the largest positive effect on reward redemptions (Exp(B) = 99.099, $p < .05$) while “Area_Porto_Mean” has the largest opposite effect (Exp(B)=0.467, $p < .05$). This means that when “Matket_Corporate_Mean” increases by 1 unit, the odds of the guest redeeming rewards increases by 99.09 times, while when “Area_Porto_Mean” increases by one unit, the odds of redeeming rewards, decreases by 0.467 times. In addition, the model reveals that “Market_Direct_Site_Mean” has also large positive effects ($Exp(B)=80.493$, $p < .05$) on reward redemptions, followed by "Market_Direct_Other_Mean" ($Exp(B)=69.464$, $p < .05$) and "Market_Direct_Contact_Center_Mean" ($Exp(B)= 69.853$, $p < .05$). Lastly, "Area_Algarve_Mean" ($Exp(B)= 3.706$, $p < .05$) and "Market_Contracted_Online_Mean" ($Exp(B)= 3.514$, $p < .05$) have also a significant positive effect on reward redemptions. On the other hand, "Type_Resort_Mean" ($Exp(B)= 0.771$, $p < .05$) and "Other_Rev_Mean" ($Exp(B)= 0.996$, $p < .05$) have small negative, yet significant, effects on reward redemptions, although the latter is reasonably smaller.

3. Discussion

This project's primary data mining goals were to identify the most important predictors of reward redemptions and patterns and common characteristics among customers who engaged in those planned behaviors. At the same time, the overall objective of this project was to provide the company with powerful insights that could aid in the design of marketing strategies and decision-making.

To begin with, descriptive statistics show that only a small percentage (2.49%) of guests redeemed rewards between 2018 and 2019, with the vast majority (66.41%) doing so only once. These findings are consistent with what the company believes to be the result of the various flaws and weaknesses identified in their loyalty program - a very low redemption rate.

According to this study, customers who redeem rewards book four times more than those who do not. This backs up the findings of Xie and Chen (2015), who found that active loyalty is related to future usage. On the other hand, these behaviors could be explained by the points pressure and rewarded behavior effects proposed by Kopalle et al. (2012) and Drèze and Nunes (2011). The former results from a conscious process in which collectors act so that they can accumulate more points in order to gain the reward, thus booking more frequently to have enough points to exchange. The latter occurs after the redemption because the customer develops beliefs in a gain or a good deal, resulting in future increased sales. This results in the willingness to extend the relationship with the hotel and continuous engagement, in other words, active loyalty. (Xie & Chen, 2014; Xie et al., 2015).

Furthermore, the analysis revealed that customers who redeem rewards spend more than those who do not, both in lodging and F&B. These findings are consistent with Palmatier et al. (2006, 2007), who claim that loyal customers are willing to pay a fair premium to the same seller.

According to the findings of the C&RT model, the most important predictor of reward redemptions is hotel booking through Hotel Group X's website. The C&RT model suggests that visitors who frequently book through the hotel group's website are more likely to redeem rewards. On the other hand, guests who always use booking channels other than the website, do not redeem rewards. Similarly, results from the Logistic Regression model suggest that the chances of redeeming rewards increase substantially when the guest books through the company’s website. Furthermore, descriptive statistics show that guests who redeem rewards are much more likely to book through the website than those who do not. Guests who do not redeem rewards, on the other hand, are more likely to book through OTA’s.

Hotel Group X encourages its loyalty program members to book through their website by offering a 10% discount on reservations and significantly more points than if they book through a travel agency. However, travel agencies, especially OTA’s, are very popular among travelers. They offer multiple alternatives, comparisons, and prices to the customers, allowing them to search by price, location, or other criteria. Moreover, these third-party players have the advantage of product bundling with air travel, car rentals, among others (O’Connor, 2021). In addition, OTA’s have their loyalty programs (Koo et al., 2020) and invest highly in advertising, offering reward incentives, and gaining the top slot engines (Feinstein, 2018). The findings of this study are consistent with Myung and Bai (2009) and O'Connor (2021), who stated that OTA’s provide hotels with a way to reach customers who are not brand loyal and purchase rooms based solely on price. In contrast, loyal customers who do not take price as the only variable tend to book through direct channels, especially the hotel’s website.

In fact, results from the LR showed that all direct channels are highly significant in the model and increase the likelihood of redeeming rewards. However, the LR model also suggests that the odds of redeeming rewards increase by 3.5 times when the customer books through an OTA. This idea might seem contradictory, but a possible explanation could be that customers have multiple memberships and take advantage of both OTA’s and Hotel’s loyalty programs. This support the findings of Xie and Chen (2015) who state that once there are no membership fees, customers are encouraged to enroll in multiple programs and shop around. Hendler et al. (2021) further state that customers might be using loyalty programs opportunistically.

Findings from the C&RT model suggest that the second most important predictor of reward redemptions is staying in Algarve. Likewise, the LR model proposes that staying in Algarve increases the chances of redeeming rewards. These results are not surprising, since the Algarve region accounted for 33% of total overnight stays in Portugal in 2019. Moreover, it accounts for 35.8% of total beds in the country and the region with the highest capacity per hotel (average of 287.9 beds per hotel) (INE, 2020, p. 30). Regarding Hotel Group X, Algarve accounts for 16% of its hotel units and 22% of total rooms. As a result, it is reasonable to conclude that these findings may have been skewed by the number of hotels in the Algarve region and the number of tourists who visit the region each year. In addition, Hotel Group X ran multiple "cash&points" promotions throughout 2019. Algarve hotels received 32% of the budget for these marketing campaigns, while city hotels received 28%. In these promotions, smaller quantities of points (from 1000 to 4000 points + cash) might be exchanged for stays in specific hotel units.

Results from the C&RT model are vague and inconclusive regarding predictions on who will actually redeem rewards. It suggests that clients who book through the website and stay in city hotels frequently redeem points and customers who make reservations through the company's website and alternate between staying in city hotels and hotel units in Algarve redeem awards. It does not, however, provide clear conclusions about how frequently these visitors stay in those locations. It is only possible to understand that such guests visited Algarve and cities at least once, which does not imply that they are preferred destinations or strong antecedents of reward redemptions.

On the other hand, C&RT model can be quite effective in predicting who will not redeem rewards. Results showed that customers who rarely use the website to make their reservations do not redeem rewards. At the same time customers who stay almost every time in city hotels and use the website to make the reservation, do not redeem rewards. Finally, customers rarely staying in Algarve and city hotels, booking through the website, end up not redeeming rewards as well. Similarly, the LR model indicates that the chances of redemption are clearly decreased when customers travel a lot to Porto or stay in resorts very often.

A major finding from the LR model was the huge significance that the corporate market segment has on the likelihood of reward redemptions. The chances of redemption are clearly increased in corporate customers. This may suggest that corporate guests – who are generally business travelers – accumulate points in their work trips (presumably staying in city hotels) and exchange the points earned for leisure trips (possibly in Algarve).

Interestingly, contrary to guests who redeem rewards, customers who do not, stay in Madeira very often. It suggests that Madeira is also a preferred destination among guests. Therefore it leads us to believe that during the promotional campaigns, customers take advantage of these initiatives, choosing the destination based on the selected hotels and their points rather than the destination per se.

4. Managerial Implications and Recommendations

Due to global competitiveness, obtaining knowledge about consumer attributes and analyzing previously recorded data to generate relationship marketing tactics and CRM strategies has become critical for hotel organizations. This study provides insights into the determinants of active loyalty in Hotel Group X's guests in the form of reward redemptions. In addition, it presents strategies for an improved management of this type of loyalty. From a marketing standpoint, these findings can aid in strategy design and decision making. It helps identifying client clusters that share similar characteristics, behaviors, and preferences.

One of the key findings of this project was the critical importance that reservations through the company's website have in active loyalty. This is a good measure of devotion, yet it was discovered that not only customers who do not redeem rewards, but also some customers who do, continue to book through OTA's predominantly. This suggests that those clients do not see the advantage in booking directly through hotel's own channels. Hotels must make sure that potential customers have as many touchpoints as possible throughout the whole user journey, when using the website, visiting social media and during ad campaigns. These touchpoints should be defined and developed in a way that let customers know how much it pays off to be loyal to the hotel group and to book directly with the hotel.

Two more discoveries will provide great value for Hotel Group X's top management. According to data, active loyal customers respond positively to point promotions for hotel units in Algarve. Madeira, on the other hand, is a popular choice among loyalty program members and potentially a profitable market. It would be worthwhile to, on one hand, keep promoting Algarve because of its obvious great acceptance and, on the other hand, launch point promotions in Madeira. Since it is such a popular destination among members, those who already participate in promotions would have an additional destination to choose from, while those who do not, but enjoy the destination, would be more likely to participate in these promotions and start engaging in the program.

Once the models become quite accurate in predicting who will not engage in redemption activities, company will be able to shift its focus and avoid wasting resources in markets where the promotions are less well-received or where customers are less engaged in the program (e.g., Porto). However, launching promotional campaigns with higher conversion rates could be one strategy to capture these customers in these markets. Unlike "cash&points" campaigns, where customers must already have points to redeem, this type of initiative would encourage customers to start accumulating points by rewarding them with more points for each euro spent.

The relevance of the corporate market segment is another important finding. Both the employee and the company that pays for the stay are customers in this segment. At the end of the day, the customer is the program's member and the one who benefits from it. However, if the paying company was also given points for each reservation, it would be a win-win situation. The company would be encouraged to make more reservations, and this important market segment would expand even further, eventually with the entry of new active loyal program members.

The findings can help to improve the practical direction for hotel management. As observed, the correct management of the large volume of data generated by clients during their stay, which is regularly collected in the CRM system, allows for a greater and better understanding of their characteristics. The ability to alter offerings in real-time depending on a customer's dynamic behavior allows the tourism industry to develop a better understanding of what they value and how loyal they are. The models developed in this study can be used with different data in the future. Because machine learning is adaptable, forecasts can alter when new data is introduced in the model, facilitating the generation of fresh, meaningful insights.

5. Limitations

The study's most significant flaw is the lack of records in the predicted class=1, or customers who redeemed rewards. Even though the dataset under study was large, with nearly two thousand customers, only a small percentage of them (2.5%) redeemed rewards. The study's power is reduced, and the margin of error is increased when the sample size is too small. Furthermore, it resulted in an unbalanced dataset, which affected the model's performance and the ability to generalize the results.

Moreover, there was a significant amount of incorrect data, primarily regarding revenues (accommodation and F&B), which was presumably due to front-office typing errors. Some reservations had large revenue amounts, which could have skewed the results.

The intervals in the nodes were too wide, resulting in inconclusive results from the C&RT model. For example, it suggested that guests who stayed in a city hotel for more than 0.021 (ranging from 0 to 1, where 0 means he never stayed and 1 means he always does) redeem rewards. However, it is impossible to determine the frequency with which he visits this type of hotels. Therefore, although the C&RT model produced an accurate decision tree, the results may have been insufficiently conclusive due to ambiguous splitting values.

Finally, as previously stated, Algarve accounted for 33% of all overnight stays in Portugal in 2019 and 22% of all Hotel Group X units. Therefore, findings may have been skewed by the number of hotels in the Algarve region and the number of tourists who visit the region each year.

6. References

Aggarwal, C. C., & Zhai, C. (2012). An Introduction to Text Mining. In *Mining Text Data* (pp. 1–10). Springer US. https://doi.org/10.1007/978-1-4614-3223-4_1

Akroush, M. N., Dahiyat, S. E., Gharaibeh, H. S., & Abu-Lail, B. N. (2011). Customer relationship management implementation. *International Journal of Commerce and Management*, *21*(2), 158–190. https://doi.org/10.1108/10569211111144355

Anshari, M., Almunawar, M. N., Lim, S. A., & Al-Mudimigh, A. (2019). Customer relationship management and big data enabled: Personalization & customization of services. *Applied Computing and Informatics*, *15*(2), 94–101. https://doi.org/10.1016/j.aci.2018.05.004

Bagozzi, R., Gopinath, M., & Nyer, P. (1999). "Impact of business intelligence and predictive analysis in big data. *Journal of the Academy of Marketing Science, 27*(2), 184–206.

Baloglu, S. (2002). Dimensions of Customer Loyalty: Separating Friends from Well Wishers. *The Cornell Hotel and Restaurant Administration Quarterly*, *43*(1), 47–59. https://doi.org/10.1177/0010880402431005

Barton J Goldenberg. (2002). *CRM Automation*. Prentice Hall Professional.

Bello-Orgaz, G., Jung, J. J., & Camacho, D. (2016). Social big data: Recent achievements and new challenges. *Information Fusion*, *28*, 49–59. https://doi.org/10.1016/j.inffus.2015.08.005

Bendapudi, N., & Leone, R. P. (2003). Psychological Implications of Customer Participation in Co-Production. *Journal of Marketing*, *67*(1), 14–28. https://doi.org/10.1509/jmkg.67.1.14.18592

Berry L L. (1983). Relationship marketing. *Emerging Perspectives on Services Marketing, American Marketing Association*, 25–38.

Bhattacharjee D, Seeley J, & Seitzman N. (2017, October 3). *Advanced analytics in hospitality*. https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/advanced-analytics-in-hospitality

Bolden D, Hadlock P, & Melker K. (2014). *Leveraging the Loyalty Margin: Rewards Programs That Work*. https://www.bcg.com/publications/2014/retail-transportation-travel-tourism-leveraging-loyalty-margin-rewards-programs-work

Boulding, W., Staelin, R., Ehret, M., & Johnston, W. J. (2005). A Customer Relationship Management Roadmap: What is Known, Potential Pitfalls, and Where to Go. *Journal of Marketing*, *69*(4), 14–18. https://doi.org/10.1509/jmkg.2005.69.4.155

Bowen, J. T., & Chen, S. (2001). The relationship between customer loyalty and customer satisfaction. *International Journal of Contemporary Hospitality Management*, *13*(5), 213–217. https://doi.org/10.1108/09596110110395893

Bowen, J. T., & Shoemaker, S. (2003). Loyalty: A Strategic Commitment. *Cornell Hotel and Restaurant Administration Quarterly*, *44*(5–6), 12–25. https://doi.org/10.1177/001088040304400505

Brown, J., Broderick, A. J., & Lee, N. (2007). Word of mouth communication within online communities: Conceptualizing the online social network. *Journal of Interactive Marketing*, *21*(3), 2–20. https://doi.org/10.1002/dir.20082

Brown T J, Mowen H, Todd D, & Licatta J. (2002). The customer orientation of service workers: personality trait determinants and effect on self and supervisor performance ratings. *Journal of Marketing Research, 39*(1), 110–119.

Burnett, S. (2020, April 27). *Redemption Rates for Loyalty Programs*. https://www.customerinsightgroup.com/loyaltyblog/loyalty-marketing/redemption-rates

Caie, P. D., Dimitriou, N., & Arandjelović, O. (2021). Precision medicine in digital pathology via image analysis and machine learning. In *Artificial Intelligence and Deep Learning in Pathology*. Elsevier. https://doi.org/10.1016/B978-0-323-67538-3.00008-7

Chang, W., Park, J. E., & Chaiy, S. (2010). How does CRM technology transform into organizational performance? A mediating role of marketing capability. *Journal of Business Research*, *63*(8), 849–855. https://doi.org/10.1016/j.jbusres.2009.07.003

Chapman P, & Clinton J. (2000). CRISP-DM 1.0: Step-by-step data mining guide. *Computer Science*.

Chaudhuri, A., & Holbrook, M. B. (2001). The Chain of Effects from Brand Trust and Brand Affect to Brand Performance: The Role of Brand Loyalty. *Journal of Marketing*, *65*(2), 81–93. https://doi.org/10.1509/jmkg.65.2.81.18255

Chawla, N. v., Japkowicz, N., & Kotcz, A. (2004). Editorial. *ACM SIGKDD Explorations Newsletter*, *6*(1), 1–6. https://doi.org/10.1145/1007730.1007733

Chen, I. J., & Popovich, K. (2003). Understanding customer relationship management (CRM). *Business Process Management Journal*, *9*(5), 672–688. https://doi.org/10.1108/14637150310496758

Chi, C. G., & Gursoy, D. (2009). Employee satisfaction, customer satisfaction, and financial performance: An empirical examination. *International Journal of Hospitality Management*, *28*(2), 245–253. https://doi.org/10.1016/j.ijhm.2008.08.003

Dalci, I., Tanis, V., & Kosan, L. (2010). Customer profitability analysis with time-driven activity-based costing: a case study in a hotel. *International Journal of Contemporary Hospitality Management*, *22*(5), 609–637. https://doi.org/10.1108/09596111011053774

Demirci Orel, F., & Kara, A. (2014). Supermarket self-checkout service quality, customer satisfaction, and loyalty: Empirical evidence from an emerging market. *Journal of Retailing and Consumer Services*, *21*(2), 118–129. https://doi.org/10.1016/j.jretconser.2013.07.002

Dick, A. S., & Basu, K. (1994). Customer Loyalty: Toward an Integrated Conceptual Framework. *Journal of the Academy of Marketing Science*, *22*(2), 99–113. https://doi.org/10.1177/0092070394222001

Diebner R, Malfara D, Neher K, Thompson M, & Vancauwenberghe M. (2021, February 24). *Prediction: The future of CX*. https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/prediction-the-future-of-cx

Dimitriadis, S., & Stevens, E. (2008). Integrated customer relationship management for service activities. *Managing Service Quality: An International Journal*, *18*(5), 496–511. https://doi.org/10.1108/09604520810898857

Donavan D T, Brown T J, & Mowen J C. (2004). Internal benefits of service-worker customer orientation: job satisfaction, commitment, and organizational citizenship behaviors. *Journal of Marketing*, *68*(1), 128–145.

Doyle, P. (1995). Marketing in the new millennium. *European Journal of Marketing*, *29*(13), 23–41. https://doi.org/10.1108/03090569510147712

Drèze, X., & Nunes, J. C. (2009). Feeling Superior: The Impact of Loyalty Program Structure on Consumers' Perceptions of Status. *Journal of Consumer Research*, *35*(6), 890–905. https://doi.org/10.1086/593946

Drèze, X., & Nunes, J. C. (2011). Recurring Goals and Learning: The Impact of Successful Reward Attainment on Purchase Behavior. *Journal of Marketing Research*, *48*(2), 268–281. https://doi.org/10.1509/jmkr.48.2.268

Dursun, A., & Caber, M. (2016). Using data mining techniques for profiling profitable hotel customers: An application of RFM analysis. *Tourism Management Perspectives*, *18*, 153–160. https://doi.org/10.1016/J.TMP.2016.03.001

Dutu C, & Halmajan H. (2011). The effect of organizational readiness on CRM and business performance. *International Journal of Computers*, *1*(2), 106–114.
Elliott, A., & Woodward, W. (2014). *IBM SPSS by Example: A Practical Guide to Statistical Data Analysis* (2nd ed.). SAGE Publications, Inc.
Evanschitzky, H., Ramaseshan, B., Woisetschläger, D. M., Richelsen, V., Blut, M., & Backhaus, C. (2012). Consequences of Customer Loyalty to the loyalty program and to the company. *Journal of the Academy of Marketing Science*, *40*(5), 625–638. https://doi.org/10.1007/s11747-011-0272-3
Feinstein E. (2018, February 23). *OTA's vs. direct hotel bookings: Which is the leading trend for 2018?* https://www.traveldailynews.com/post/otas-vs-direct-hotel-bookings-which-is-the-leading-trend-for-2018
Fenech C, & Perkins B. (2017). *Customer loyalty: A relationship, not just a scheme*. https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/consumer-business/deloitte-uk-consumer-review-customer-loyalty.pdf
Fox T, & Stead S. (2001). Customer Relationship Management: Delivering the Benefits. *CRM (UK) and SECOR Consulting*.
Frow, P. E., & Payne, A. F. (2009). Customer Relationship Management: A Strategic Perspective. *Journal of Business Market Management*, *3*(1), 7–27. https://doi.org/10.1007/s12087-008-0035-8
Gantz, J., & Reinsel, D. (2011). *Extracting Value from Chaos*. http://www.emc.com/digital_universe.
Garbarino, E., & Johnson, M. S. (1999). The Different Roles of Satisfaction, Trust, and Commitment in Customer Relationships. *Journal of Marketing*, *63*(2), 70–87. https://doi.org/10.2307/1251946
Garrido-Moreno, A., & Padilla-Meléndez, A. (2011). Analyzing the impact of knowledge management on CRM success: The mediating effects of organizational factors. *International Journal of Information Management*, *31*(5), 437–444. https://doi.org/10.1016/j.ijinfomgt.2011.01.002
Gebert, H., Geib, M., Kolbe, L., & Brenner, W. (2003). Knowledge-enabled customer relationship management: integrating customer relationship management and knowledge management concepts[1]. *Journal of Knowledge Management*, *7*(5), 107–123. https://doi.org/10.1108/13673270310505421
Glassoff S, Gross M, Reichheld A, & Sebag Y. (2020). *Airlines & Hospitality: The values of loyalty in a crisis* . https://www.deloittedigital.com/content/dam/deloittedigital/us/documents/blog/blog-20200618-airlines-hospitality-covid.pdf
González-Serrano, L., Talón-Ballestero, P., Muñoz-Romero, S., Soguero-Ruiz, C., & Rojo-Álvarez, J. L. (2019). Entropic Statistical Description of Big Data Quality in Hotel Customer Relationship Management. *Entropy 2019, Vol. 21, Page 419*, *21*(4), 419. https://doi.org/10.3390/E21040419
González-Serrano, L., Talón-Ballestero, P., Muñoz-Romero, S., Soguero-Ruiz, C., & Rojo-Álvarez, J. L. (2020). A Big Data Approach to Customer Relationship Management Strategy in Hospitality Using Multiple Correspondence Domain Description. *Applied Sciences*, *11*(1), 256. https://doi.org/10.3390/app11010256
Grönroos, C. (1994). From Marketing Mix to Relationship Marketing. *Management Decision*, *32*(2), 4–20. https://doi.org/10.1108/00251749410054774
Hamid H. (2009). Toward unfolding CRM implementation in Pakistan: a case study", . *17th European Conference on Information Systems*, 496–511.
Hand D J, Mannila H, & Smyth P. (2001). *Principles of Data Mining*. MIT Press.

Hansotia, B. (2002). Gearing up for CRM: Antecedents to successful implementation. *Journal of Database Marketing & Customer Strategy Management*, *10*(2), 121–132. https://doi.org/10.1057/palgrave.jdm.3240103

Hendler, F., & Latour, K. A. (2008). A Qualitative Analysis of Slot Clubs as Drivers of Casino Loyalty. *Cornell Hospitality Quarterly*, *49*(2), 105–121. https://doi.org/10.1177/1938965508316017

Hendler, F., LaTour, K. A., & Cotte, J. (2021). Temporal Orientation and Customer Loyalty Programs. *Cornell Hospitality Quarterly*. https://doi.org/10.1177/19389655211008413

Higgins M. (2020, January 6). *How data and analytics are changing the hotels industry*. https://pwc.blogs.com/industry_perspectives/2020/01/how-data-and-analytics-are-changing-the-hotels-industry.html

Hong-kit Yim F, Anderson R E, & Swaminathan S. (2004). Customer Relationship Management: Its Dimensions and Effect on Customer Outcomes. *Journal of Personal Selling & Sales Management*, *24*(4), 263–278.

IBM. (n.d.). *IBM SPSS Modeler 18.2.2 Modeling Nodes* (p. 83).

IBM. (2021). *Partition node (SPSS Modeler) - IBM Cloud Pak for Data*. https://dataplatform.cloud.ibm.com/docs/content/wsd/nodes/partition.html

INE. (2020). *Estatísticas do Turismo – 2019*.

INE. (2021). *Estatísticas do Turismo - 2020*.

Iriana, R., & Buttle, F. (2007). Strategic, Operational, and Analytical Customer Relationship Management. *Journal of Relationship Marketing*, *5*(4), 23–42. https://doi.org/10.1300/J366v05n04_03

Jennings S, Murali R, Giorgio P, & Goggin S C. (2014). *Winning the race for guest loyalty When frequent travelers choose a favorite program, they aren't the only ones who reap rewards.* https://www2.deloitte.com/content/dam/Deloitte/tr/Documents/consumer-business/winning-the-race-hotel-loyalty-pov-final.pdf

Kandampully, J., & Suhartanto, D. (2000). Customer loyalty in the hotel industry: the role of customer satisfaction and image. *International Journal of Contemporary Hospitality Management*, *12*(6), 346–351. https://doi.org/10.1108/09596110010342559

Karakostas, B., Kardaras, D., & Papathanassiou, E. (2005). The state of CRM adoption by the financial services in the UK: an empirical investigation. *Information & Management*, *42*(6), 853–863. https://doi.org/10.1016/j.im.2004.08.006

(Karen) Xie, L., & Chen, C.-C. (2014). Hotel loyalty programs: how valuable is valuable enough? *International Journal of Contemporary Hospitality Management*, *26*(1), 107–129. https://doi.org/10.1108/IJCHM-08-2012-0145

Keh H, & Lee Y. (2006). Do reward programs build loyalty for services?The moderating effect of satisfaction on type and timing of rewards. *Journal of Retailing*, *82*(2), 127–146. https://doi.org/10.1016/j.jretai.2006.02.004

Khodakarami, F., & Chan, Y. E. (2014). Exploring the role of customer relationship management (CRM) systems in customer knowledge creation. *Information & Management*, *51*(1), 27–42. https://doi.org/10.1016/j.im.2013.09.001

Kim, W. G., & Cha, Y. (2002). Antecedents and consequences of relationship quality in hotel industry. *International Journal of Hospitality Management*, *21*(4), 321–338. https://doi.org/10.1016/S0278-4319(02)00011-7

King, S. F., & Burgess, T. F. (2008). Understanding success and failure in customer relationship management. *Industrial Marketing Management*, *37*(4), 421–431. https://doi.org/10.1016/j.indmarman.2007.02.005

Koo, B., Yu, J., & Han, H. (2020). The role of loyalty programs in boosting hotel guest loyalty: Impact of switching barriers. *International Journal of Hospitality Management*, *84*, 102328. https://doi.org/10.1016/j.ijhm.2019.102328

Kopalle, P. K., Sun, Y., Neslin, S. A., Sun, B., & Swaminathan, V. (2012). The Joint Sales Impact of Frequency Reward and Customer Tier Components of Loyalty Programs. *Marketing Science*, *31*(2), 195–368. https://doi.org/10.1287/mksc.1110.0687

Kotler P. (1992). It's Time for Total Marketing. *Business Week Advance Briefs*, *2*, 1–21.

Kumar, V., & Reinartz, W. (2018). Loyalty Programs: Design and Effectiveness. In *Customer Relationship Management* (pp. 179–205). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-55381-7_10

Lam, S. Y., Shankar, V., Erramilli, M. K., & Murthy, B. (2004). Customer Value, Satisfaction, Loyalty, and Switching Costs: An Illustration From a Business-to-Business Service Context. *Journal of the Academy of Marketing Science*, *32*(3), 293. https://doi.org/10.1177/0092070304263330

Landau, S., & Everitt, B. (2004). *A Handbook of Statistical Analyses Using SPSS*. CRC Press LLC.

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. In *META Group Research Note* (p. 70).

Lee, J., Lee, J., & Feick, L. (2001). The impact of switching costs on the customer satisfaction-loyalty link: mobile phone service in France. *Journal of Services Marketing*, *15*(1), 35–48. https://doi.org/10.1108/08876040110381463

Lee, J.-S., Tsang, N., & Pan, S. (2015). Examining the differential effects of social and economic rewards in a hotel loyalty program. *International Journal of Hospitality Management*, *49*, 17–27. https://doi.org/10.1016/j.ijhm.2015.05.003

Lee, M., Cai, Y. (Maggie), DeFranco, A., & Lee, J. (2020). Exploring influential factors affecting guest satisfaction. *Journal of Hospitality and Tourism Technology*, *11*(1), 137–153. https://doi.org/10.1108/JHTT-07-2018-0054

Leenheer, J., van Heerde, H. J., Bijmolt, T. H. A., & Smidts, A. (2007). Do loyalty programs really enhance behavioral loyalty? An empirical analysis accounting for self-selecting members. *International Journal of Research in Marketing*, *24*(1), 31–47. https://doi.org/10.1016/j.ijresmar.2006.10.005

Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management*, *68*, 301–323. https://doi.org/10.1016/j.tourman.2018.03.009

Liebowitz, J. (2016). *Beyond Knowledge Management: What Every Leader Should Know*. Auerbach Publications.

Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, *29*(3), 458–468. https://doi.org/10.1016/j.tourman.2007.05.011

Liu H Y. (2007). Development of a framework for customer relationship management (CRM) in the banking industry. *International Journal of Management*, *24*(1), 15–32.

Liu-Thompkins, Y., & Tam, L. (2013). Not All Repeat Customers Are the Same: Designing Effective Cross-Selling Promotion on the Basis of Attitudinal Loyalty and Habit. *Journal of Marketing*, *77*(5), 21–36. https://doi.org/10.1509/jm.11.0508

Lo, A. S., Stalcup, L. D., & Lee, A. (2010). Customer relationship management for hotels in Hong Kong. *International Journal of Contemporary Hospitality Management*, *22*(2), 139–159. https://doi.org/10.1108/09596111011018151

Maheshwari, A. (2014). *Business Intelligence and Data Mining*. Business Expert Press.

Main A, Stephan A, Arnason B, & Sedivy P. (2020). *COVID-19: Maintaining customer loyalty and trust during times of uncertainty*. https://www2.deloitte.com/content/dam/Deloitte/ca/Documents/finance/ca-en-customer-pov-aoda.pdf

Mariani, M. (2020). Big Data and analytics in tourism and hospitality: a perspective article. *Tourism Review*, *75*(1), 299–303. https://doi.org/10.1108/TR-06-2019-0259
McCall, M., & Voorhees, C. (2010). The Drivers of Loyalty Program Success. *Cornell Hospitality Quarterly*, *51*(1), 35–42. https://doi.org/10.1177/1938965509355395
Mechinda, P., & Patterson, P. G. (2011). The impact of service climate and service provider personality on employees' customer-oriented behavior in a high-contact setting. *Journal of Services Marketing*, *25*(2), 101–113. https://doi.org/10.1108/08876041111119822
Mendoza, L. E., Marius, A., Pérez, M., & Grimán, A. C. (2007). Critical success factors for a customer relationship management strategy. *Information and Software Technology*, *49*(8), 913–945. https://doi.org/10.1016/j.infsof.2006.10.003
Meuter, M. L., Ostrom, A. L., Roundtree, R. I., & Bitner, M. J. (2000). Self-Service Technologies: Understanding Customer Satisfaction with Technology-Based Service Encounters. *Journal of Marketing*, *64*(3), 50–54. https://doi.org/10.1509/jmkg.64.3.50.18024
Miles, M., Huberman, A., & Saldana, J. (2013). *Qualitative Data Analysis*. Sage.
Minghetti, V. (2003). Building Customer Value in the Hospitality Industry: Towards the Definition of a Customer Centric Information System. *Information Technology & Tourism*, *6*(2), 141–152. https://doi.org/10.3727/109830503773048246
Mohammad A, Basri R, & Shaharuddin T. (2013). Assessing the influence of customer relationship management (CRM) dimensions on organization performance. *Journal of Hospitality and Tourism Technology*, *4*(3), 228–247. https://doi.org/10.1108/JHTT-01-2013-0002
Moisen, G. (2008). Classification and regression trees. In *Encyclopedia of Ecology* (Vol. 8, pp. 582–588). Elsevier.
Morgan, R. M., & Hunt, S. D. (1994). The Commitment-Trust Theory of Relationship Marketing. *Journal of Marketing*, *58*(3), 20. https://doi.org/10.2307/1252308
Moriarty, J., Jones, R., Rowley, J., & Kupiec-Teahan, B. (2008). Marketing in small hotels: a qualitative study. *Marketing Intelligence & Planning*, *26*(3), 293–315. https://doi.org/10.1108/02634500810871348
Myung, E., Li, L., & Bai, B. (2009). Managing the Distribution Channel Relationship With E-Wholesalers: Hotel Operators' Perspective. *Journal of Hospitality Marketing & Management*, *18*(8), 811–828. https://doi.org/10.1080/19368620903235837
Nastasoiu, A., Bendle, N. T., Bagga, C. K., Vandenbosch, M., & Navarro, S. (2021). Separating customer heterogeneity, points pressure and rewarded behavior to assess a retail loyalty program. *Journal of the Academy of Marketing Science*. https://doi.org/10.1007/s11747-021-00782-2
Ngai E W T. (2005). Customer relationship management research (1992-2002): an academic literature review and classification. *Marketing Intelligence Planning*, *23*(6), 582–605.
O'Connor, P. (2021). Loyalty Programs and Direct Website Performance: An Empirical Analysis of Global Hotel Brands. In *Information and Communication Technologies in Tourism 2021* (pp. 150–161). Springer International Publishing. https://doi.org/10.1007/978-3-030-65785-7_13
Oliver, R. L. (1999). Whence Consumer Loyalty? *Journal of Marketing*, *63*(4_suppl1), 33–44. https://doi.org/10.1177/00222429990634s105
O'Mahony, G. B., Sophonsiri, S., & Turner, L. W. (2013). The impact of the antecedents of relationship development on Thai and Australian resort hotels guests. *International Journal of Hospitality Management*, *34*, 214–226. https://doi.org/10.1016/j.ijhm.2013.03.009

Palmatier, R. W., Dant, R. P., Grewal, D., & Evans, K. R. (2006). Factors Influencing the Effectiveness of Relationship Marketing: A Meta-Analysis. *Journal of Marketing*, *70*(4), 136–153. https://doi.org/10.1509/jmkg.70.4.136

Palmatier, R. W., Scheer, L. K., & Steenkamp, J.-B. E. M. (2007). Customer Loyalty to Whom? Managing the Benefits and Risks of Salesperson-Owned Loyalty. *Journal of Marketing Research*, *44*(2), 185–199. https://doi.org/10.1509/jmkr.44.2.185

Park, C., & Lee, T. M. (2009). Information direction, website reputation and eWOM effect: A moderating role of product type. *Journal of Business Research*, *62*(1), 61–67. https://doi.org/10.1016/j.jbusres.2007.11.017

Parvatiyar A, & Sheth J N. (2001). Customer Relationship Management: Emerging Practice, Process, and Discipline. *Journal of Economic and Social Research*, *3*(2), 1–34. https://www.researchgate.net/publication/312458264_Customer_relationship_management_Emerging_practice_process_and_discipline

Payne, A., & Frow, P. (2005). A Strategic Framework for Customer Relationship Management. *Journal of Marketing*, *69*(4), 167–176. https://doi.org/10.1509/jmkg.2005.69.4.167

Payne, A., & Frow, P. (2006). Customer Relationship Management: from Strategy to Implementation. *Journal of Marketing Management*, *22*(1–2), 135–168. https://doi.org/10.1362/026725706776022272

Peppers D, & Rogers M. (1993). *The One to One Future*.

Pesonen, J., Komppula, R., & Murphy, J. (2019). Plastic loyalty – Investigating loyalty card programs for a Finnish hotel chain. *Tourism Management*, *73*, 115–122. https://doi.org/10.1016/j.tourman.2019.01.023

Petrick, J. F. (2004). Are loyal visitors desired visitors? *Tourism Management*, *25*(4), 463–470. https://doi.org/10.1016/S0261-5177(03)00116-X

Piskar, F., & Faganel, A. (2009). A Successful CRM Implementation Project in a Service Company: Case Study. *Organizacija*, *42*(5), 199–208. https://doi.org/10.2478/v10051-009-0017-y

Plessis M D, & Boon J. (2004). Knowledge management in e-business and customer relationship management: South African case study findings. *International Journal of Knowledge Management*, *24*, 73–86.

Rahimi, R. (2017). Customer relationship management (people, process and technology) and organisational culture in hotels. *International Journal of Contemporary Hospitality Management*, *29*(5), 1380–1402. https://doi.org/10.1108/IJCHM-10-2015-0617

Rahimi, R., & Gunlu, E. (2016). Implementing Customer Relationship Management (CRM) in hotel industry from organizational culture perspective. *International Journal of Contemporary Hospitality Management*, *28*(1), 89–112. https://doi.org/10.1108/IJCHM-04-2014-0176

Ramos, C. M. Q., Martins, D. J., Serra, F., Lam, R., Cardoso, P. J. S., Correia, M. B., & Rodrigues, J. M. F. (2017). Framework for a Hospitality Big Data Warehouse. *International Journal of Information Systems in the Service Sector*, *9*(2), 27–45. https://doi.org/10.4018/IJISSS.2017040102

Reinartz, W., Krafft, M., & Hoyer, W. D. (2004). The Customer Relationship Management Process: Its Measurement and Impact on Performance. *Journal of Marketing Research*, *41*(3), 293–305. https://doi.org/10.1509/jmkr.41.3.293.35991

Rivera, M. A. (2020). Big data research in hospitality: From streetlight empiricism research to theory laden research. *International Journal of Hospitality Management*, *86*, 102447. https://doi.org/10.1016/j.ijhm.2019.102447

Roberts M L, Liu R R, & Hazard K. (2005). Strategy, technology and organisational alignment: key components of CRM success. *The Journal of Database Marketing and Customer Strategy Management*, *12*(4), 315–326.

Roger J Baran, Robert J Galka, & Daniel P Strunk. (2008). *Principles of Customer Relationship Management*. Australia South-Western 2008.

Rothschild, M. L., & Gaidis, W. C. (1981). Behavioral Learning Theory: Its Relevance to Marketing and Promotions. *Journal of Marketing*, *45*(2), 70–78. https://doi.org/10.1177/002224298104500207

Ryals, L., & Payne, A. (2001). Customer relationship management in financial services: towards information-enabled relationship marketing. *Journal of Strategic Marketing*, *9*(1), 3–27. https://doi.org/10.1080/713775725

Sarmaniotis, C., Assimakopoulos, C., & Papaioannou, E. (2013). Successful implementation of CRM in luxury hotels: determinants and measurements. *EuroMed Journal of Business*, *8*(2), 134–153. https://doi.org/10.1108/EMJB-06-2013-0031

Segel L H, Auerbach P, & Segev I. (2013). *The Power of Points: Strategies for making loyalty programs work*. https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/Marketing%20and%20Sales/Our%20Insights/The%20power%20of%20points%20Strategies%20for%20making%20loyalty%20programs%20work/Loyalty_market_power_of_points.pdf

Sharma, S. (2020). Big Data Analytics for Customer Relationship Management: A Systematic Review and Research Agenda. *Proceedings of the International Conference on Advances in Computing and Data Sciences*, , 430–448.

Sharp, B., & Sharp, A. (1997). Loyalty programs and their impact on repeat-purchase loyalty patterns. *International Journal of Research in Marketing*, *14*(5), 473–486. https://doi.org/10.1016/S0167-8116(97)00022-0

Sheth J N, & Parvatiyar A. (1994). *Relationship marketing : theory, methods, and applications*.

Shmueli, G., Patel, N., & Bruce, P. (2016). *Data Mining for Business Analytics: Concepts, Techniques, and Applications with XLMiner*. John Wiley and Sons.

Shoemaker, S., & Lewis, R. C. (1999). Customer loyalty: the future of hospitality marketing. *International Journal of Hospitality Management*, *18*(4), 345–370. https://doi.org/10.1016/S0278-4319(99)00042-0

Sigala, M. (2005). Integrating customer relationship management in hotel operations: managerial and operational implications. *International Journal of Hospitality Management*, *24*(3), 391–413. https://doi.org/10.1016/J.IJHM.2004.08.008

Sin, L. Y. M., Tse, A. C. B., & Yim, F. H. K. (2005). CRM: conceptualization and scale development. *European Journal of Marketing*, *39*(11/12), 1264–1290. https://doi.org/10.1108/03090560510623253

Sirirak, S., Islam, N., & Ba Khang, D. (2011). Does ICT adoption enhance hotel performance? *Journal of Hospitality and Tourism Technology*, *2*(1), 34–49. https://doi.org/10.1108/17579881111112403

Smart Vision Europe. (2021). *What is the CRISP-DM methodology*. https://www.sv-europe.com/crisp-dm-methodology/

Smith, A., & Sparks, L. (2009). “It’s nice to get a wee treat if you’ve had a bad week”: Consumer motivations in retail loyalty scheme points redemption. *Journal of Business Research*, *62*(5), 542–547. https://doi.org/10.1016/j.jbusres.2008.06.013

Smith, A., & Sparks, L. (2020). Reward Redemption Behaviour in Retail Loyalty Schemes. *British Journal of Management*, *20*(2), 204–218. https://doi.org/10.1111/j.1467-8551.2008.00561.x

Stringfellow, A., Nie, W., & Bowen, D. E. (2004). CRM: Profiting from understanding customer needs. *Business Horizons*, *47*(5), 45–52. https://doi.org/10.1016/j.bushor.2004.07.008

Sudhir H. Kale. (2004). CRM Failure and the Seven Deadly Sins. *Marketing Management*, *13*, 42–46.

"Sunny" Hu, H.-H., Huang, C.-T., & Chen, P.-T. (2010). Do reward programs truly build loyalty for lodging industry? *International Journal of Hospitality Management*, *29*(1), 128–135. https://doi.org/10.1016/j.ijhm.2009.07.002

Tajeddini, K. (2010). Effect of customer orientation and entrepreneurial orientation on innovativeness: Evidence from the hotel industry in Switzerland. *Tourism Management*, *31*(2), 221–231. https://doi.org/10.1016/j.tourman.2009.02.013

Talón-Ballestero, P., González-Serrano, L., Soguero-Ruiz, C., Muñoz-Romero, S., & Rojo-Álvarez, J. L. (2018). Using big data from Customer Relationship Management information systems to determine the client profile in the hotel sector. *Tourism Management*, *68*, 187–197. https://doi.org/10.1016/j.tourman.2018.03.017

Tan, A. (1999). Text mining: the state of the art and the challenges". *Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases*, 65–70.

Tanford, S., Raab, C., & Kim, Y.-S. (2011). The Influence of Reward Program Membership and Commitment on Hotel Loyalty. *Journal of Hospitality & Tourism Research*, *35*(3), 279–307. https://doi.org/10.1177/1096348010382236

Taylor, G. A., & Neslin, S. A. (2005). The current and future sales impact of a retail frequency reward program. *Journal of Retailing*, *81*(4), 293–305. https://doi.org/10.1016/j.jretai.2004.11.004

Teo, T. S. H., Devadoss, P., & Pan, S. L. (2006). Towards a holistic perspective of customer relationship management (CRM) implementation: A case study of the Housing and Development Board, Singapore. *Decision Support Systems*, *42*(3), 1613–1627. https://doi.org/10.1016/j.dss.2006.01.007

Tuominen, M., Rajala, A., & Möller, K. (2004). Market-driving versus market-driven: Divergent roles of market orientation in business relationships. *Industrial Marketing Management*, *33*(3), 207–217. https://doi.org/10.1016/j.indmarman.2003.10.010

Turismo de Portugal. (2021, May 7). *Visão Geral*. http://www.turismodeportugal.pt/pt/Turismo_Portugal/visao_geral/Paginas/default.aspx

Ul-Haq, R. (1994). Relationship Marketing—Bringing Quality, Customer Service and Marketing Together. M. Christopher, A. Payne and D. Ballantyne, Butterworth-Heinemann, Oxford, 1991, 204 pp, ISBN 0 750602589, price £25.00. *Strategic Change*, *3*(2), 119–120. https://doi.org/10.1002/jsc.4240030208

UNWTO. (n.d.). *Tourism – an economic and social phenomenon*. Retrieved September 5, 2021, from https://www.unwto.org/why-tourism

UNWTO. (2020, January 19). *World Tourism Barometer Nº18 January 2020*. https://www.unwto.org/world-tourism-barometer-n18-january-2020

UNWTO. (2021, January 28). *2020: Worst Year In Tourism History With 1 Billion Fewer International Arrivals*. https://www.unwto.org/news/2020-worst-year-in-tourism-history-with-1-billion-fewer-international-arrivals

Vaeztehrani, A., Modarres, M., & Aref, S. (2015). Developing an integrated revenue management and customer relationship management approach in the hotel industry. *Journal of Revenue and Pricing Management*, *14*(2), 97–119. https://doi.org/10.1057/rpm.2014.22

Verhoef, P. C. (2003). Understanding the Effect of Customer Relationship Management Efforts on Customer Retention and Customer Share Development. *Journal of Marketing*, *67*(4), 30–45. https://doi.org/10.1509/jmkg.67.4.30.18685

Weissenberg A, Katz A, & Narula A. (2013). *A Restoration in Hotel Loyalty: Developing a blueprint for reinventing loyalty programs*. https://www2.deloitte.com/content/dam/Deloitte/us/Documents/consumer-business/us-thl-customer-loyalty-pov.pdf

Weissenberg A, & Langford G. (2018). *Moving the global travel industry forward*. https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Consumer-Business/deloitte-wttc-moving-global-travel-industry-forward.pdf

Woolan R, Davis P, de Angelis F, & Quiring K. (2017). *Seeing beyond the loyalty illusion: it's time you invest more wisely*. https://www.accenture.com/_acnmedia/pdf-43/accenture-strategy-gcpr-customer-loyalty.pdf

World Travel & Tourism Council. (2020). *City Travel & Tourism Impact 2019*. https://wttc.org/Portals/0/Documents/Reports/2019/City%20Travel%20and%20Tourism%20Impact%20Graphics%20Report%20Dec%202019.pdf?ver=2021-02-25-201320-033

Wu, S.-I., & Chen, J.-H. (2012). Comparison between hotels and motels using CRM effect model – An empirical study in Taiwan. *International Journal of Hospitality Management*, *31*(4), 1254–1263. https://doi.org/10.1016/j.ijhm.2012.03.005

Wu, S.-I., & Lu, C.-L. (2012). The relationship between CRM, RM, and business performance: A study of the hotel industry in Taiwan. *International Journal of Hospitality Management*, *31*(1), 276–285. https://doi.org/10.1016/j.ijhm.2011.06.012

Xiang, Z., Schwartz, Z., Gerdes, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, *44*, 120–130. https://doi.org/10.1016/j.ijhm.2014.10.013

Xie, K. L., & Chen, C.-C. (2013). Progress in Loyalty Program Research: Facts, Debates, and Future Research. *Journal of Hospitality Marketing & Management*, *22*(5), 463–489. https://doi.org/10.1080/19368623.2012.686148

Xie, K. L., Xiong, L., Chen, C.-C., & Hu, C. (2015). Understanding Active Loyalty Behavior in Hotel Reward Programs Through Customers' Switching Costs and Perceived Program Value. *Journal of Travel & Tourism Marketing*, *32*(3), 308–324. https://doi.org/10.1080/10548408.2014.896767

Xiong, L., King, C., & Hu, C. (2014). Where is the love? *International Journal of Contemporary Hospitality Management*, *26*(4), 275–292. https://doi.org/10.1108/IJCHM-03-2013-0141

Yau, O. H. M., Lee, J. S. Y., Chow, R. P. M., Sin, L. Y. M., & Tse, A. C. B. (2000). Relationship marketing the Chinese way. *Business Horizons*, *43*(1), 16–24. https://doi.org/10.1016/S0007-6813(00)87383-8

Yoo, M., & Bai, B. (2013). Customer loyalty marketing research: A comparative approach between hospitality and business journals. *International Journal of Hospitality Management*, *33*, 166–177. https://doi.org/10.1016/j.ijhm.2012.07.009

Zablah, A. R., Bellenger, D. N., & Johnston, W. J. (2004). An evaluation of divergent perspectives on customer relationship management: Towards a common understanding of an emerging phenomenon. *Industrial Marketing Management*, *33*(6), 475–489. https://doi.org/10.1016/j.indmarman.2004.01.006

Zeithaml, V. A. (2000). Service Quality, Profitability, and the Economic Worth of Customers: What We Know and What We Need to Learn. *Journal of the Academy of Marketing Science*, *28*(1), 67–85. https://doi.org/10.1177/0092070300281007

Zhang, J. Q., Craciun, G., & Shin, D. (2010). When does electronic word-of-mouth matter? A study of consumer product reviews. *Journal of Business Research*, *63*(12), 1336–1341. https://doi.org/10.1016/j.jbusres.2009.12.011

Zhu Z, & Nakata C. (2007). Reexamining the link between customer orientation and business performance: the role of information systems. *Journal of Marketing Theory and Practice, 15*(3), 187–203.