

Leveraging Data Warehousing and Mining to Improve Customer Satisfaction in the Hospitality Sector

¹*Rida Farheen Mohamed Sameeullah, Noorunnisa Sulthan Alavudeen,*

²*Dr. Siddhaling Urolagin,*

^{1,2}*Department of Computer Science Engineering and Information Technology,*

²*Middlesex University Dubai, Dubai Campus, Dubai Internet City, PO Box 500697, Dubai*

Abstract

In today's data-driven economy, the hospitality industry faces increasing pressure to personalize guest experiences and maintain high levels of customer satisfaction. This paper explores how integrating data warehousing and data mining techniques can provide actionable insights to enhance service delivery and guest satisfaction in the hospitality sector. The study begins with an overview of current challenges and data utilization trends within the industry, highlighting the role of structured data consolidation and analytical tools in strategic decision-making. As a practical case study, La Veranda Hotel in Larnaca, Cyprus, is examined using a two-phased experimental setup: data warehouse design using a star schema for customer-related metrics and subsequent data mining through OLAP operations and association rule mining via Weka. Findings reveal key satisfaction trends based on guest demographics, booking patterns, and room preferences, demonstrating how hospitality providers can leverage intelligent data systems to optimize guest experience. This study found that 95% of highly satisfied guests stayed fewer than five nights, over 45% gave a perfect satisfaction score, and room type, guest origin (notably UK and Cyprus), and peak months (June, Sept, Oct) significantly influenced satisfaction outcomes.

Keywords

Hospitality industry, customer satisfaction, data warehousing, data mining, Weka, business intelligence, star schema, ETL process, tourism technology, data-driven decision-making

1. Introduction

The hospitality industry is undergoing a data-driven transformation, with hotels increasingly seeking to personalize guest experiences and improve operational efficiency through digital technologies (Anwar, Deliana, and Suyamto, 2024). Central to this shift is the ability to collect, integrate, and analyze large volumes of structured and unstructured data to inform customer-centric decision-making (Alimi, Oladele, and Mia, 2024). La Veranda de Larnaca Hotel, a boutique hotel in Cyprus, serves as an ideal case study to examine how data warehousing and mining techniques can enhance customer satisfaction. The diverse data generated from guest bookings, feedback, room preferences, and third-party platforms presents a valuable opportunity to derive actionable insights—if processed effectively (Zarezadeh, Rastegar, and Xiang, 2022).

This paper explores how a hybrid data architecture, utilizing Amazon Redshift and Amazon S3 (Esther, 2024), can support the development of a robust data warehouse for La Veranda Hotel. Analytical methods such as OLAP, Tableau dashboards, and association rule mining via the Apriori algorithm in Weka are employed to uncover satisfaction trends linked to demographics, booking patterns, and room choices. Section 2 presents a literature survey covering recent trends in customer satisfaction analytics. Section 3 explains the methodology, including the star schema design and ETL pipeline. Section 4 outlines the experimental setup, while Section 5 presents the results, highlighting high-confidence patterns from the Apriori analysis. The paper concludes in Section 6 with key managerial insights and recommendations for enhancing service delivery and guest loyalty through data-driven decision-making.

2. Literature Survey

Customer satisfaction in the hospitality industry is a pivotal determinant of business success, rooted in the consistent delivery of high-quality service that exceeds guest expectations. It is both an emotional and evaluative response shaped by every interaction a customer has with the hotel, from the initial check-in to the final farewell. In this context, positive guest relationships are strategic assets that not only drive return visits but also generate positive word-of-mouth and long-term profitability (Dominici, G. and Rosa, G., 2010). Measuring satisfaction, therefore, requires a multidimensional approach, combining guest feedback, data analytics, and employee insights, to anticipate needs and enhance the guest experience meaningfully. (Ali, B., Othman, B.J., Gardi, B. and Ahmed, S.A., 2021). Guest behavior analytics enables hyper-personalized services by analyzing historical booking data, spending habits, and preferences to tailor offerings at an individual level, such as preferred room settings or personalized amenities. Predictive analytics further refines this by optimizing dynamic pricing, staffing, and operational efficiency based on real-time and historical demand patterns. Such insights not only improve the immediate guest experience but also contribute to long-term loyalty through personalized loyalty programs and targeted marketing campaigns that align with guest behavior trends. Importantly, real-time behavioral tracking allows hotels to identify and respond to negative feedback proactively, thereby preserving brand reputation. (Bhinder, H.S., 2022).

Business Intelligence (BI) tools allow hospitality companies to collect, store, and process data from multiple sources into meaningful insights. By using BI, companies gain a comprehensive view of customer behavior, market trends, and operational performance (Madyatmadja, E.D., Adiba, C.N.A., et al, 2021). BI enhances the accuracy of decision-making by organizing large datasets into actionable reports and dashboards, thereby increasing cost efficiency and reducing the time needed to derive strategic insights. Additionally, BI helps identify high-value customers, understand the factors driving their loyalty, and segment future customers with similar potential. This supports not only personalized service delivery but also improved customer retention strategies. In the hospitality sector, big data analytics plays a crucial role in uncovering guest preferences, behaviors, and trends by analyzing interactions and feedback from platforms like Facebook, Instagram, and Twitter. Through this analysis, businesses can identify patterns that are not immediately visible, enabling them to offer personalized content, tailored marketing strategies, and enhanced services that create emotional connections and loyalty. Predictive analytics allows hotels to anticipate customer needs and provide proactive recommendations, thereby improving overall guest satisfaction. With the integration of Big Data, hotels can customize the guest journey by offering personalized experiences based on historical and real-time data. This personalization is powered by targeted advertising, dynamic pricing models, and AI-enabled tools such as chatbots and smart assistants. For instance, Marriott Hotels utilizes data from guest interactions and external events (like concerts) to optimize pricing and improve guest experiences, leading to increased revenue and brand loyalty

(Tymoshchenko, D., 2024).

In the evolving landscape of the hospitality industry, data-driven strategies are becoming central to enhancing customer satisfaction and operational efficiency. However, a significant gap exists between large-scale and medium or small-scale hotels leverage data analytics. Large hotel chains have embraced data as a strategic asset, integrating advanced analytics across customer experience, marketing, and operations. Through loyalty programs and guest profiles, these hotels gather detailed insights on customer preferences, such as room types, check-in habits, and amenities usage, to deliver highly personalized experiences. This personalization extends to tailored marketing campaigns and dynamic offers that align with individual guest behavior, significantly improving customer retention and brand loyalty. (Renascence.io., 2024). In contrast, medium and low-scale hotels often struggle to adopt such data-centric approaches due to a variety of constraints. Hotels typically operate with limited financial and human capital. Investing in sophisticated data management tools, analytics platforms, and skilled personnel may be financially unfeasible, especially when immediate survival or day-to-day operations take priority. Many smaller hotels maintain traditional, experience-based decision-making models. Shifting to a data-driven culture requires not just technological adoption but a significant change in mindset, something that is often met with skepticism or inertia among staff. Some hotel operators underestimate the tangible value that data can bring. Prior successes without using data may reinforce the belief that analytics are unnecessary, despite the shifting expectations of modern, tech-savvy guests (Johansson, A., 2024). As a result, while larger hotels continue to optimize operations and customer experiences using data, many smaller players are left behind. This divide highlights the need for increased awareness, scalable solutions, and possibly collaborative platforms that can help medium-scale hotels bridge the gap and unlock the full potential of customer and operational data. As the hospitality industry increasingly generates both traditional structured data and large-scale unstructured big data, it becomes essential to adopt robust data warehousing and mining practices to manage the intricate flow of information across systems.

3. Methodology

To address the question of how La Veranda Hotel, a medium - scale hotel can use data systems to enhance guest satisfaction and loyalty, a two-pronged systematic methodology was applied. A star schema data model (Kimball and Ross) was designed to analyze hotel guest satisfaction as shown below in Fig. 1. At the centre is the GuestSatisfaction fact table, which records key metrics such as satisfaction score, sentiment score, staff service rating, overall rating, and stay duration. This fact table connects to five surrounding dimension tables that provide descriptive attributes for deeper analysis. The Guest Dimension captures demographic information including guest name, country, segment, loyalty status, and amenities used, offering insights into guest behaviour and preferences. The Review Dimension contains feedback-related attributes such as sentiment, category, staff service rating, and overall rating, allowing for textual and quantitative review analysis. The Room Dimension details characteristics like room type, view, floor, size, and whether a balcony is available, enabling segmentation of satisfaction based on room features. The Booking Dimension provides information on how the booking was made, including the source, website platform, booking status, and check-in/check-out dates. Lastly, the Time Dimension includes temporal elements such as date, day, month, year, season, promotion periods, and local events, supporting trend and seasonality analysis. Together, these dimensions offer a holistic framework for understanding and analysing guest satisfaction in a hotel setting, integrating demographic, behavioural, temporal, and service-level data.

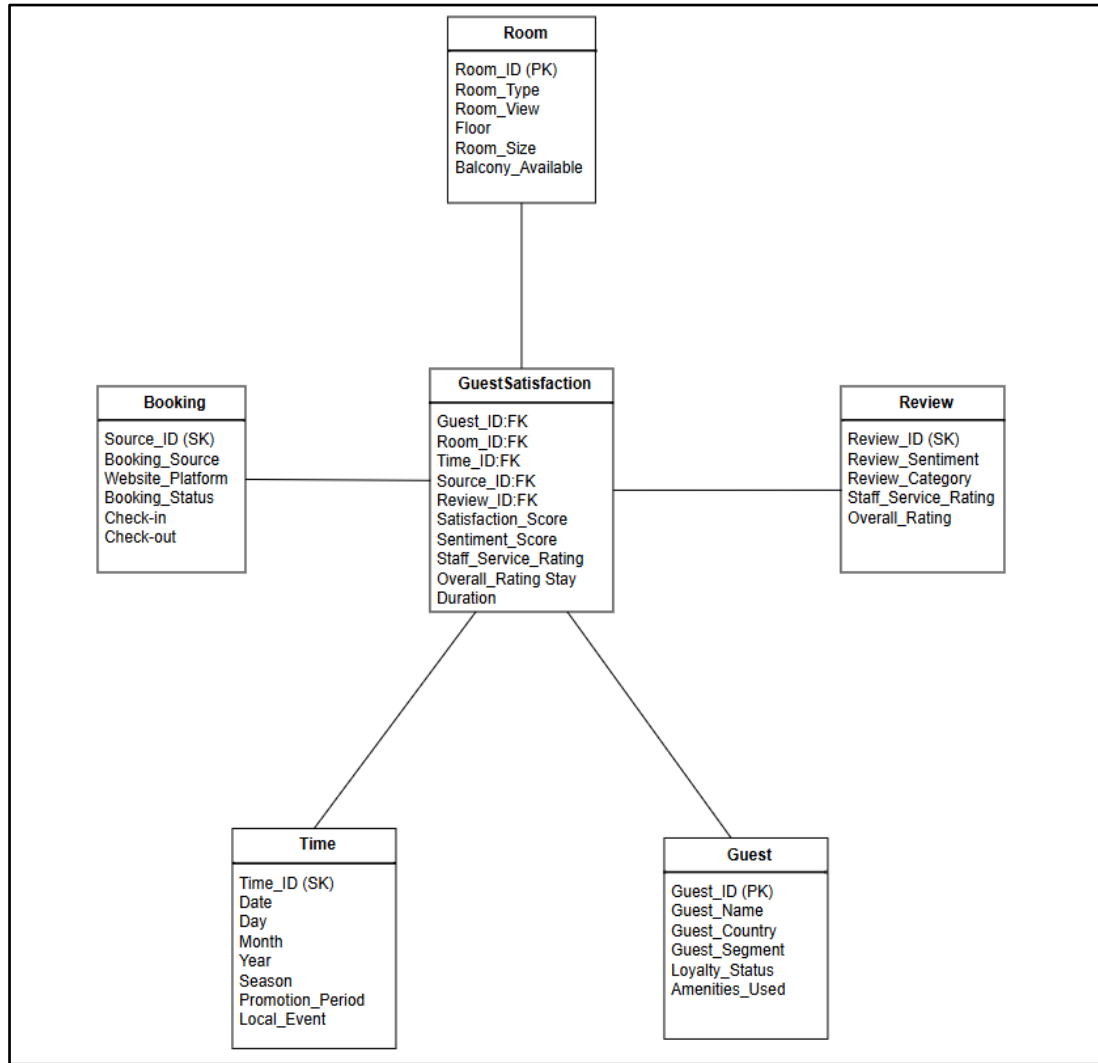


Figure 1. Star Schema for La Vernada Hotel

The data pipeline follows a clear Extract-Transform-Load (ETL) workflow in Fig. 2. The data processing for this study involved extracting both internal and external data sources. Internal data, including booking records, customer relationship management (CRM) information, and property management system (PMS) logs, were combined with external data such as online reviews, local event schedules, and weather conditions to provide a comprehensive dataset. During the transformation phase, data cleaning was performed, and continuous numeric variables were converted into nominal categories, for example, satisfaction scores were grouped into categories like GroupedScore, and length of stay was classified into NightsDuration category.

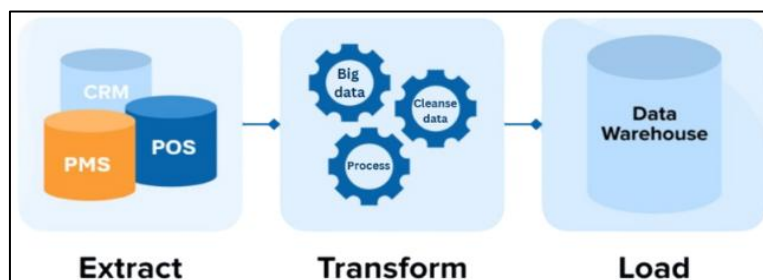


Figure 2. ETL (Extract, Transform, Load) Process Flow

Amazon Redshift is a fully managed, cloud-based data warehouse that supports complex analytical queries on large datasets using a columnar storage model, making it ideal for OLAP-style operations (Stancu, 2014). Amazon S3 (Simple Storage Service) provides durable, scalable object storage, well-suited for unstructured data like customer reviews. Finally, the structured data was loaded into Amazon Redshift for efficient querying and analysis, while unstructured data, including textual reviews, was stored in Amazon S3 to enable future sentiment analysis. This approach aligns with best practices in hospitality analytics, which emphasize integrating diverse data sources and leveraging cloud-based data warehousing for scalable analytics. (Ghosh and Sen, 2023)

4. Experimental Setup

The experimental setup involves a two-phased approach. In the first phase, a data warehouse is designed for La Veranda Hotel to consolidate data and support decision-making. Data is extracted from internal sources, including guest profiles, stay details, and booking information, as well as external sources like third-party booking details and local event schedules. This data undergoes an ETL process to ensure quality and consistency before being loaded into a star schema. The star schema comprises a central Guest Satisfaction fact table and dimension tables for Guest, Review, Room, Booking, and Time. To improve performance and query efficiency, indexing and data partitioning strategies are also applied during the data warehouse design. Metadata is documented to support long-term maintenance and facilitate data governance. OLAP techniques are considered for multidimensional analysis (IBM, 2021). The second phase focuses on analyzing the dataset that contains customer information for La Veranda Hotel (Hatab, 2023). The initial stage involves data cleaning, which includes generating unique Guest_IDs, handling inconsistencies, correcting geographical and date data, addressing missing values, and refining column values. Data normalization techniques were also applied to ensure consistency across numerical fields such as room rates and stay durations. Exploratory data analysis (EDA) helped uncover hidden patterns and correlations in guest behavior. Following cleaning, the data is used to analyze guest demographics, room preferences, satisfaction levels, and trends using Tableau Dashboards and Weka.

5. Experimental Results

The analysis of La Veranda Hotel's customer satisfaction data reveals key insights into guest behavior and preferences. Most guests originate from the UK, Cyprus, and Palestine, with double rooms, particularly those with balconies, being the most popular choice. Additionally, since the hotel mainly caters to families and larger groups and does not offer single rooms, Solo Travellers have little choice but to opt for double rooms. From Fig. 3 below, several graphs illustrate insights derived from the data of La Veranda Hotel. Tableau was employed for this analysis by uploading the dataset into the platform, and the resulting visualizations highlight these significant patterns.

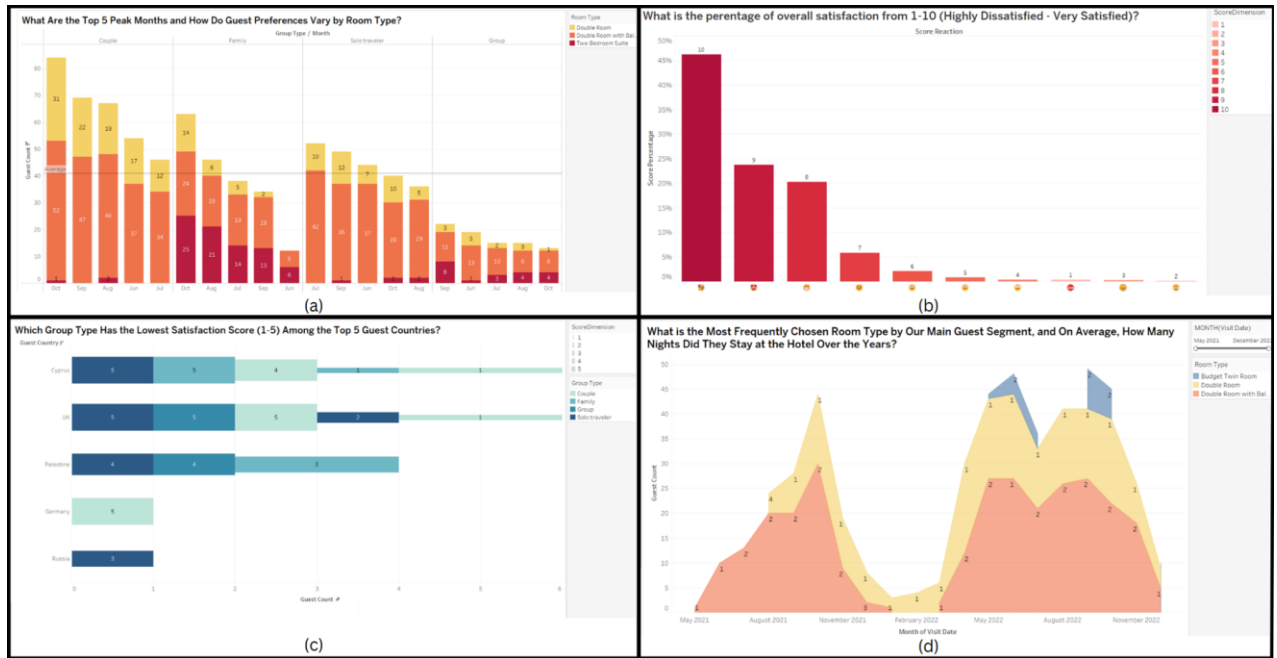


Figure 3. Insights from the hotel dataset visualized using Tableau: (a) Room Type Preference & Length of Stay, (b) Overall Satisfaction (1–10), (c) Lowest Satisfaction by Group Type and Country, (d) Peak Months and Room Preferences by Group Type.

Couples, the main guest segment, typically stay 1–2 nights and favor the Double Room with Balcony, particularly during the peak months of June, September, and October. Overall guest satisfaction is high, with over 45% awarding a perfect 10 rating and only 5% rating their stay poorly. Guests from Cyprus and the UK, especially couples and solo travelers, report the highest dissatisfaction, while families and groups tend to be more satisfied. Families show a strong preference for 2-Bedroom Suites, particularly during peak periods in August and October. Booking volumes decrease from December to March, especially among couples, reflecting a clear seasonal pattern in demand. These insights underscore the importance of targeted marketing strategies and tailored service improvements to enhance guest experiences throughout the year.

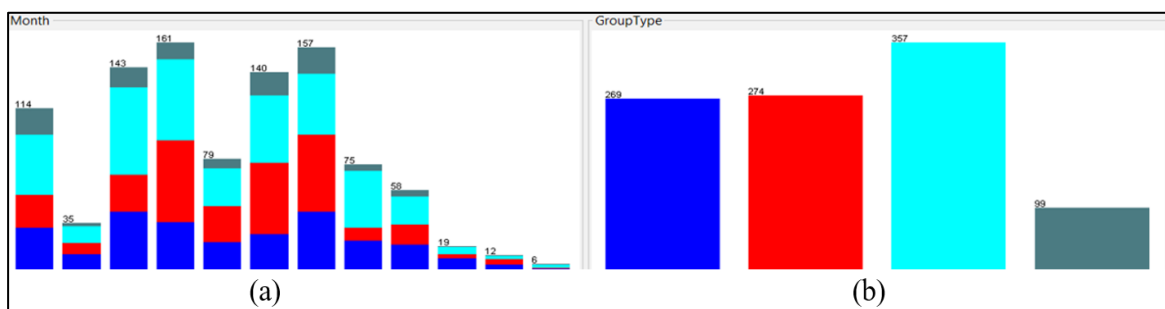


Figure 4. Distribution of Instances by (a) Month and (b) Group Type Visualized in WEKA

The Fig. 4 above displays two key visualizations from Weka’s “Visualize All” tool, showcasing the distribution of instances across Month and GroupType, with color-coded segments representing four distinct group types: Solo Traveller, Couple, Family, and Group. In the Month panel, the bookings are spread throughout the year, with peak activity observed in months like June, July, and August, suggesting a seasonal trend in hotel visits. Each bar is stacked with different group type colours, indicating the diversity of travellers throughout the year. Notably, family and couple bookings dominate the summer months. The GroupType panel shows the frequency distribution of each traveller type. It is evident that Groups and Couples make up most

bookings, while solo travellers and families are fewer in number but still significant, confirming class balance suitable for association rule mining.

This visualization validates that the dataset is rich and balanced across key categorical variables. Prior to this, preprocessing steps were undertaken to prepare the data for Apriori-based association rule mining. This involved converting numerical satisfaction scores into categories (e.g., *Highly Satisfied* = 9–10), segmenting night durations into binary categories (<5 and >5), and enriching the dataset with attributes like GroupedCountry, NightsDuration, and GroupType. These enhancements enabled the extraction of meaningful and interpretable association rules from the dataset as shown below in Table 1. For instance, one rule showed that 95.3% of guests who were highly satisfied stayed fewer than 5 nights. Another rule indicated that Asian guests had a higher likelihood of being highly satisfied in 2022, unlike their European counterparts.

Table 1. Apriori algorithm results for a selected iteration

Rule No.	Antecedent(s)	Consequent(s)	Confidence	Lift	Leverage	Conviction
1	ScoreNominal=Ten	GroupedScore= Highly Satisfied	1	1.36	0.13	129.41
2	NightsDuration<5, ScoreNominal=Ten	GroupedScore= Highly Satisfied	1	1.36	0.12	123.02
3	YearNominal=Twenty Two, ScoreNominal=Ten	GroupedScore= Highly Satisfied	1	1.36	0.09	90.8
4	NightsDuration<5, YearNominal=Twenty Two, ScoreNominal=Ten	GroupedScore= Highly Satisfied	1	1.36	0.09	87.34
5	YearNominal=Twenty Two	NightsDuration<5	0.97	1.01	0.01	1.26
6	GroupedScore=Highly Satisfied, YearNominal=Twenty Two	NightsDuration<5	0.96	1.01	0.01	1.16
7	YearNominal=Twenty Two, ScoreNominal=Ten	NightsDuration<5	0.96	1.01	0.01	1.07
8	GroupedScore=Highly Satisfied, YearNominal=Twenty Two, ScoreNominal=Ten	NightsDuration<5	0.96	1.01	0.01	1.07
9	YearNominal=Twenty Two, ScoreNominal=Ten	GroupedScore= Highly Satisfied, NightsDuration<5	0.96	1.37	0.09	7.31
10	GroupedScore=Highly Satisfied	NightsDuration<5	0.95	1	0	0.92

These rules have high support and confidence, suggesting actionable insights, such as tailoring short-stay packages or improving services for European guests. These valuable insights were derived from data on just a single boutique hotel. For hotel chains with multiple branches or those serving high volumes of guests year-round, applying data mining techniques like Apriori at scale can lead to significant operational improvements and enhanced customer satisfaction. By continuously analyzing guest behavior and feedback, hotels can adapt dynamically to customer needs, streamline services, and stay competitive in the evolving hospitality landscape. (Yu, 2025)

6. Conclusion

This research highlights the transformative impact of leveraging data warehousing and mining to improve customer satisfaction in the hospitality sector. By designing a star schema-based data warehouse and applying the Apriori algorithm using Weka, the study uncovered actionable insights into guest preferences, booking behaviours, and satisfaction patterns at La Veranda Hotel. Among the key findings, the Apriori algorithm revealed that guests who stayed fewer than five nights and gave a top score were also highly satisfied with 100% confidence (Leverage = 0.12). These high-confidence, high-leverage rules highlight the strong relationship between short stay durations and satisfaction, providing a solid basis for tailoring short-stay experience packages and promotional strategies.

The findings demonstrate that even a single-location hotel can derive immense strategic value from structured data analysis, revealing trends such as shorter stays correlating with higher satisfaction and geographic variation in guest experiences. The methodology of integrating OLAP operations with association rule mining, proves effective for supporting guest-centric service enhancement. As the hospitality industry continues to embrace digital transformation, the use of hybrid architectures (e.g., AWS S3 and Redshift) and intelligent mining tools will be essential for scaling these benefits across larger hotel chains. (Daisy, 2025). Hence, data-driven decision-making enables hotels to refine offerings, personalize services, and foster long-term guest loyalty. The insights gained from this case study at La Veranda Hotel provide a scalable blueprint for other hospitality providers seeking to enhance operational efficiency and elevate customer satisfaction through modern data practices.

References

1. Anwar, F.A., Deliana, D. and Suyamto Suyamto (2024). Digital Transformation in the Hospitality Industry: Improving Efficiency and Guest Experience. *International Journal of Management Science and Information Technology*, [online] 4(2), pp.428–437. doi: <https://doi.org/10.35870/ijmsit.v4i2.320>.
2. Alimi, A., Oladele, S. and Mia, J. (2024). *Data-Driven Decision-Making: Challenges and Opportunities in Business Analytic*. [online] ResearchGate. Available at: https://www.researchgate.net/publication/387365368_Data-Driven_Decision-Making_Challenges_and_Opportunities_in_Business_Analytic.
3. Zarezadeh, Zara & Rastegar, Raymond & Xiang, Zheng. (2022). Big Data Analytics and Hotel Guest Experience: A Critical Analysis of the Literature. *International Journal of Contemporary Hospitality Management*. 34. 10.1108/IJCHM-10-2021-1293.
4. Hatab, M. (2023). *Hotel Reviews Booking.com*. [online] Kaggle.com. Available at: <https://www.kaggle.com/datasets/michelhatab/hotel-reviews-bookingcom>
5. Esther, Dorcas. (2024). Data Lake Architecture for Scalable Analytics on AWS S3 and Redshift.
6. Dominici, G. and Rosa, G. (2010). Customer Satisfaction in the Hotel Industry: A Case Study from Sicily. *International Journal of Marketing Studies*, 2(2), pp.3–3. [online] Available at: https://www.researchgate.net/publication/47619084_Customer_Satisfaction_in_the_Hotel_Industry_A_Case_Study_from_Sicily#references
7. Ali, B., Othman, B.J., Gardi, B. and Ahmed, S.A. (2021). Hotel Service Quality: The Impact of Service Quality on Customer Satisfaction in Hospitality. *International Journal of Engineering Business and Management*, [online] Available at: https://www.researchgate.net/profile/Bayad-Ali/publication/351780048_Hotel_Service_Quality_The_Impact_of_Service_Quality_on_Customer_Satisfaction_in_Hospitality/links/60a9741592851ca9dcdad96a/Hotel-Service-Quality-The-Impact-of-Service-Quality-on-Customer-Satisfaction-in-Hospitality.pdf
8. Bhinder, H.S. (2022). Emerging Technology Trends Adopted in Hotel Guest Cycle for Enhancing Guest Stay Experiences. *Recent Trends in Tourism and Hospitality*, pp.18–29. [online] Available at: https://www.researchgate.net/profile/Mahesh-Bairwa-2/publication/363925145_RECENT_TRENDS_IN_TOURISM_HOSPITALITY/links/63356405ff870c55cee7e0e4/RECENT-TRENDS-IN-TOURISM-HOSPITALITY.pdf#page=36

9. Tymoshchenko, D. (2024). How Big Data and Analytics Are Transforming the Hospitality Industry. [online] *Acropolium*. Available at: <https://acropolium.com/blog/big-data-in-the-hospitality-industry/>.
10. Madyatmadja, E.D., Adiba, C.N.A., Sembiring, D.J.M. and Pristinella, D. (2021). The Positive Impact of Implementation Business Intelligence and Big Data in Hospitality and Tourism Sector. *International Journal of Emerging Technology and Advanced Engineering*, 11(6), pp.59–71. [online] Available at: https://www.researchgate.net/publication/352462853_The_Positive_Impact_of_Implementation_Business_Intelligence_and_Big_Data_in_Hospitality_and_Tourism_Sectorurl.
11. *Renascence.io*. (2024). How Hilton Enhances Customer Experience (CX) with Digital Innovations in Hospitality. [online] *Renascence.io*. Available at: <https://www.renascence.io/journal/how-hilton-enhances-customer-experience-cx-with-digital-innovations-in-hospitality>
12. Kimball, R. and Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. 3rd ed. Wiley.
13. Johansson, A. (2024). Data Dilemma: Is the Hospitality Industry Lagging Behind? [online] *Demand Calendar*. Available at: <https://www.demandcalendar.com/blog/data-dilemma-is-the-hospitality-industry-lagging-behind>.
14. Stancu, A.-M.R. (2014). OLAP systems - Solutions for multidimensional data analysis. *ResearchGate*, [online] 15, pp.75–80. Available at: https://www.researchgate.net/publication/297328274_OLAP_systems_-_Solutions_for_multidimensional_data_analysis.
15. Ghosh, M. and Sen, S. (2023). *CLOUD-BASED DATA ANALYTICS PLATFORM IN THE HOSPITALITY INDUSTRY: A COMPREHENSIVE ANALYSIS AND FUTURE PROSPECTS*. [online] ResearchGate. Available at: https://www.researchgate.net/publication/375755932_CLOUD-BASED_DATA_ANALYTICS_PLATFORM_IN_THE_HOSPITALITY_INDUSTRY_A_COMPREHENSIVE_ANALYSIS_AND_FUTURE_PROSPECTS.
16. IBM (2021). What is OLAP? IBM Knowledge Center. [online] Available at: <https://www.ibm.com/think/topics/olap>
17. Yu, R.-Z. (2025). The Impact of Digital Transformation in the Hospitality Industry on Tourism Consumption Experience. *Global Economic Perspectives*, [online] 3(1). doi: <https://doi.org/10.37155/2972-4813-0301-4>.
18. Daisy, E. (2025). *Scaling Intelligent Decision Engines Across Complex Enterprise System*. [online] ResearchGate. Available at: https://www.researchgate.net/publication/390998460_Scaling_Intelligent_Decision_Engines_Across_Complex_Enterprise_System.
19. Esther, Dorcas. (2024). Data Lake Architecture for Scalable Analytics on AWS S3 and Redshift.