Multidimensional Analysis of Booking Data in Hospitality Industry Using Data Warehousing Techniques

Lam Mai[*], Duc-Hien Nguyen

The University of Danang, Vietnam - Korea University of Information and Communication Technology, Danang, Vietnam

[*]**Email:** mlam@vku.udn.vn

Abstract

Understanding customer behaviors is essential for optimizing pricing strategies, enhancing guest experiences, and effectively meeting demand in the hospitality industry. This study presents the development of a data warehouse system designed to analyze hotel booking behaviors. Using ETL processes, reservation data from diverse sources is consolidated and standardized to enable comprehensive analysis. Then, multidimensional analyses of booking frequency and transaction value reveal key customer preferences and behavioral patterns. A real-world dataset comprising 119,390 records spanning from July 1, 2015, to August 31, 2017, was utilized to validate the system. Multidimensional analyses revealed that 70% of bookings occurred during peak seasons, with transaction values averaging 25% higher compared to off-peak periods. Additionally, customers who booked via direct channels displayed a 20% higher retention rate. The results validate the proposed system's capability to provide actionable intelligence, driving effective business strategies and supporting predictive modeling in the hospitality industry.

Keywords

Introduction

In the era of data-driven decision-making, the hospitality industry, particularly hotels, faces mounting pressure to leverage data analytics to remain competitive and responsive to market demands. Despite the growing body of research on customer behavior and pricing strategies in the hospitality industry, there remains a gap in the integration of these elements into a unified data-driven framework. Previous studies have typically focused on individual aspects, such as booking patterns or customer preferences, but few have explored how these areas interconnect to optimize pricing and service delivery across various hotel segments. This study addresses this gap by developing a comprehensive data warehouse system that integrates performance metrics, historical booking data, and customer behavior to provide actionable insights.

Recently, studies by Badgujar (2023) on booking patterns primarily focuses on enhancing customer segmentation, while studies on customer preferences by (Dolnicar, 2022) often overlook

the influence of seasonal demand. By synthesizing these approaches, this study provides a holistic view to enhance hotel management decision-making. The key areas analyzed include:

Performance Analysis: Evaluating booking data, room preferences, and stay durations to identify factors affecting hotel performance, such as high-demand seasons and pricing strategies.

Historical Booking Analysis: Analyzing booking history to uncover seasonal trends and demand patterns for better strategy formulation.

Customer Analysis: Proposing a customer management module for collecting and analyzing customer demographics, booking behaviors, and preferences to support personalized promotions and services.

The insights derived from this research are expected to contribute to the development of tailored strategies, enabling hotels to achieve better market alignment, improved profitability, and sustained competitive advantage.

Historical data analysis plays a key role in understanding booking behaviors and predicting future demand. The study by (Talluri et al., 2014) have explored revenue management strategies using historical booking data, emphasizing metrics like Revenue per Available Room (RevPAR). Addtionally, (Al-Debei et al., 2016) showed that analyzing cancellation patterns and no-show rates could significantly improve operational efficiency and minimize revenue loss.

Research on customer behavior has focused on understanding demographic characteristics, booking preferences, and spending patterns. Studies by (Dolnicar et al., 2022) explored methods for customer segmentation in the hospitality industry, identifying distinct market segments such as business travelers, leisure tourists, and families. Recent studies have leveraged machine learning techniques to refine segmentation and enhance personalization (Sahu et al, 2022). These findings highlight the importance of detailed customer profiles in crafting targeted marketing strategies and improving guest satisfaction.

Data warehousing has been extensively explored as a solution for integrating and managing large datasets. (Kimball et al., 2009) emphasized the role of ETL (Extract, Transform, Load) processes in ensuring data quality and consistency for analytical purposes. In the hospitality industry, ETL methods have been employed to consolidate transactional data from diverse sources, such as online booking platforms, property management systems, and customer relationship management. OLAP (Online Analytical Processing) tools that is leveraging this multidimensional data, enable complex analyses and provide valuable insights, helping businesses optimize strategies and improve decision-making processes. Ivanov et al. (2019) had highlighted its application in analyzing high-demand periods, preferred room types, and the impact of promotional campaigns.

While substantial research has been conducted on data warehousing and analytics in the hospitality industry, gaps remain in integrating these methodologies into comprehensive decision-support systems. Few studies (Wardhani et al., 2024; Taufik et al., 2021) have explored the combined use of ETL processes, OLAP tools, and customer segmentation techniques in the context of hotel performance analysis. Data-driven methodologies in optimizing hotel operations and enhancing customer satisfaction. However, there is a need for more integrated approaches that combine advanced analytics with practical applications. This study seeks to address these gaps by

developing a data warehouse-based framework for hotel performance analysis, incorporating multidimensional analytics, customer segmentation, and real-time booking tracking.

Table 1. Hotel booking data in tabular format

#	Field Name	Description	Datatype
1	is_canceled	Indicates if the booking was canceled or not	Integer
2	arrival_date_year	Year of the arrival data	Integer
3	arrival_date_month	Month of the arrival data (12 categories: January to December)	Categorial
4	arrival_date_week_number	Week number of the arrival date	Integer
5	stays_in_weekend_nights	Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay	Integer
6	country	Country of origin. Categories are represented in the ISO 3155–3:2013 format	Object
7	previous_cancellations	Number of previous bookings that were cancelled by the customer prior to the current booking	Integer
8	reservation_status	Reservation last status, assuming one of three categories: Canceled; Check-Out; No-Show	Object

Methodology

Data Preparation
The data preparation phase is essential for developing a data warehouse to analyze customer booking behaviors in the hotel industry including Information Requirements Analysis identifies key stakeholders, including hotel managers and marketing teams. Goals such as analyzing booking trends, customer demographics, and seasonal demand are established through consultations and operational reviews, guiding the selection of analytical dimensions. And Data Source Identification that consolidates data from: Property Management Systems (PMS): Room bookings, customer profiles, and stay details; Booking Platforms: Online travel agency (OTA) and direct booking data, including times, preferences, and promotions; Revenue Systems: Metrics on rates, discounts, and revenue; CRM Systems: Customer contact information, booking history, and feedback.

Hotel Booking Demand Datasets was proposed by (Antonio et al, 2019). that provides a wide-ranging collection of 119,390 real-world data records during the historical period between July 1, 2015, and August 31, 2017. The dataset offers valuable perceptions into hotel booking demand, capturing the real-life dynamics of the hospitality industry, with information on various factors such as booking dates, customer behavior, and hotel characteristics, personal details related to hotel or customer identification have been completely removed to protect privacy. Table 1 illustrated hotel booking data in a tabular format, consisting of eight key fields that provide valuable insights into booking patterns and customer behaviors.

Figure 1 illustrates the proposed architecture of our data warehouse, which consists of four key components: data preparation, data integration, data analysis, and data visualization. A collaborative, data-driven approach ensures accurate integration of these sources into a multidimensional schema. Critical dimensions, such as booking status, customer segments, and room types, are defined. ETL (Extract, Transform, Load) processes is mentioned in the next part that cleans and standardizes the data for analysis.
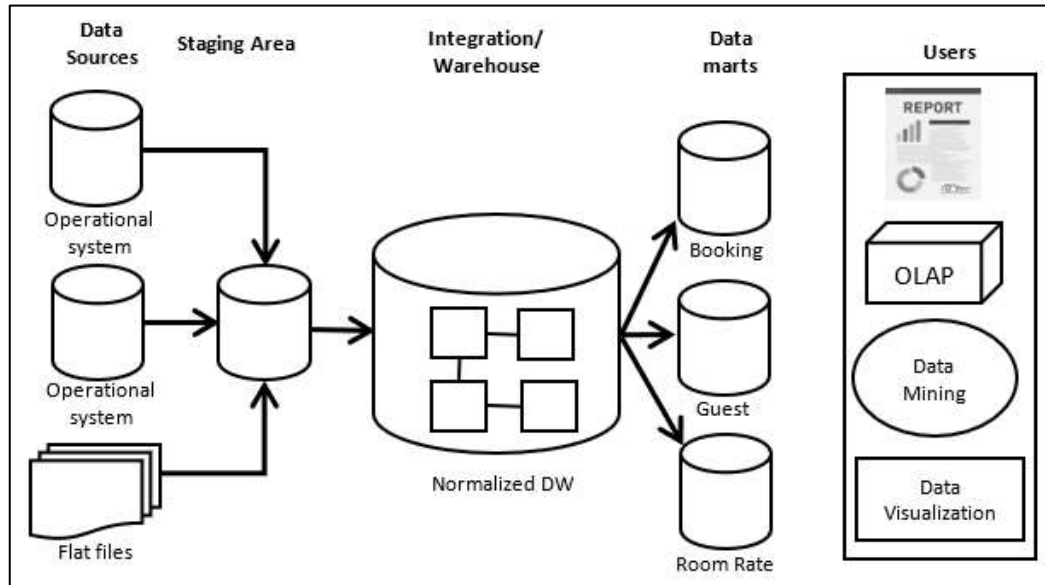

Figure 1. Our data warehouse proposed architecture

Data Integration

In the ETL process, data is first extracted from multiple sources, including the hotel reservation system, customer databases, and external market data. The extraction process involves pulling raw data files (CSV) and querying relational databases. In the transformation phase, we clean the data by handling missing values, standardizing date formats, and combining tables from different sources to ensure consistency. We also perform data aggregation for different customer segments (e.g., by location, booking frequency). Finally, the transformed data is loaded into our data warehouse, where it is stored in a structured format to support OLAP analysis. Figure 2 shows the denormalized multidimensional facts star schema from our proposed method, which is designed to simplify the structure of the data warehouse by organizing it into fact tables and dimension tables.
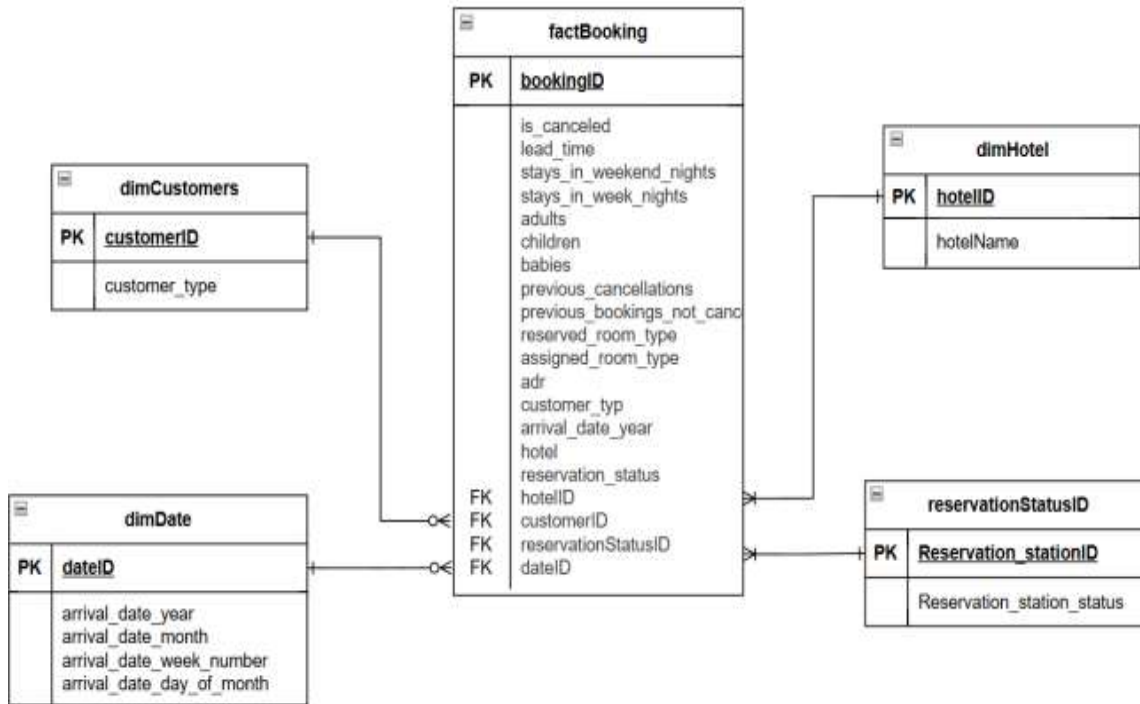
Figure 2. Denormalized multidimensional facts star schema

Data Analysis

In the analytics process, the OLAP server operates as the core component in the middle tier of the data warehouse architecture. Data aggregation is structured into OLAP cubes, where specific dimension tables are associated with corresponding fact tables, and the fact tables provide the measurable data points. Multidimensional analysis enables users to perform operations such as drilling down into granular details or drilling across related dimensions, achieved by querying the fact tables. And this phase is crucial for delivering user-centric visualizations, including detailed reports, graphical representations, and interactive dashboards, ensuring that stakeholders can easily interpret and utilize the data for informed decision-making.

Data Visualization

The results of the hotel data analysis are visualized through a variety of chart types, leveraging OLAP cube functionalities to query the data warehouse. Bar charts, line charts, and pie charts are utilized to represent different analytical findings and reports. While the dashboard acts as the central user interface for the hotel data warehouse system, presenting a comprehensive view of analytical results. It aggregates multiple charts and metrics into a single, visually accessible platform, allowing users to explore a wide range of information, including booking trends, customer demographics, revenue metrics, and occupancy rates, across various dimensions such as time, location, and room type.

Results and Discussion

Hotel Booking Trends Analysis

The number of hotel reservations demonstrates continuous fluctuations throughout 2017. During the early months of the year, particularly from April to August, the number of bookings remains relatively low. However, as summer approaches, reservations show a noticeable increase, especially in June and July. In the autumn months, particularly October and November, there is a significant rise in bookings, this study found that 70% of bookings occurred during peak seasons, with an average transaction value 25% higher than off-peak periods which could be attributed to holidays or special events taking place during this period. Figure 3 shows that in 2017, the number of hotel reservations was affected by the seasons and special events. Higher peaks in the summer and fall months may be due to holidays, tourist seasons or special events. This information can help hoteliers to plan better to manage capacity and provide services throughout the year.
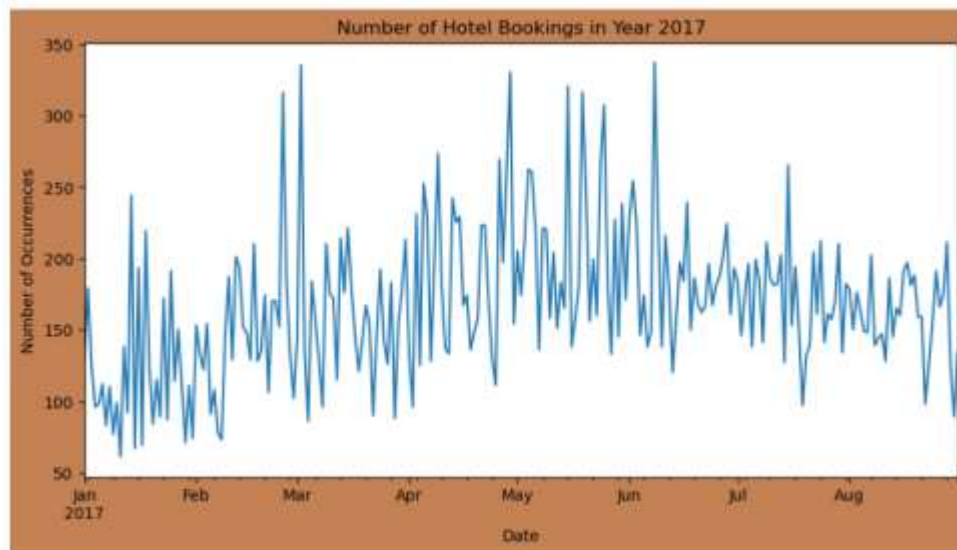

Figure 3. Impact of seasons and special events on hotel reservations in 2017

Booking Channels Analysis

Figure 4 shows the number of reservations through different reservation methods or channels in hotels. This segmentation provides a clear understanding of regional and demographic influences on booking behavior. Online travel agencies are the leading booking channel, requiring hotels to prioritize partnerships with these platforms. However, traditional offline bookings and group reservations still hold a significant market share. Direct bookings remain important, particularly for travelers seeking direct engagement with hotels. Customer retention rates were found to be 20% higher for bookings made through direct channels. Strategic management of corporate and airline bookings can also yield targeted benefits despite their smaller scale.
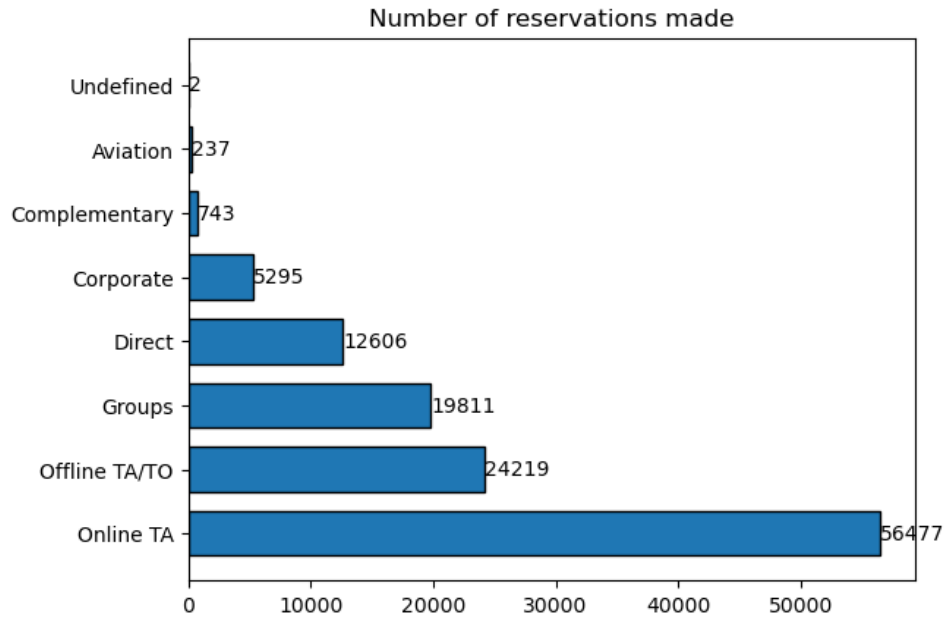
Figure 4. Distribution of hotel booking across different booking channels

Booking Cancellations and No-Shows

The analysis of hotel booking data reveals two key trends: the majority of reservations lead to guest stays, while cancellations constitute a notable proportion. Out of a total of 40,060 reservations, 28,938 rooms were not canceled, indicating that most bookings successfully reached the final stage. However, 11,122 reservations were canceled, showing that cancellations remain a considerable factor in the booking process as illustrated Figure 5.

The ratio of canceled to non-canceled rooms stands at approximately 38.4%, meaning for every 100 rooms successfully booked, nearly 38 are canceled. This translates to the number of non-canceled bookings being 2.6 times higher than cancellations. While the high success rate highlights the hotel's ability to retain most bookings, the significant cancellation rate suggests room for improvement. Analyzing these trends further can help uncover customer satisfaction issues, identify common reasons for cancellations, and enhance reservation processes to minimize losses and boost operational efficiency.
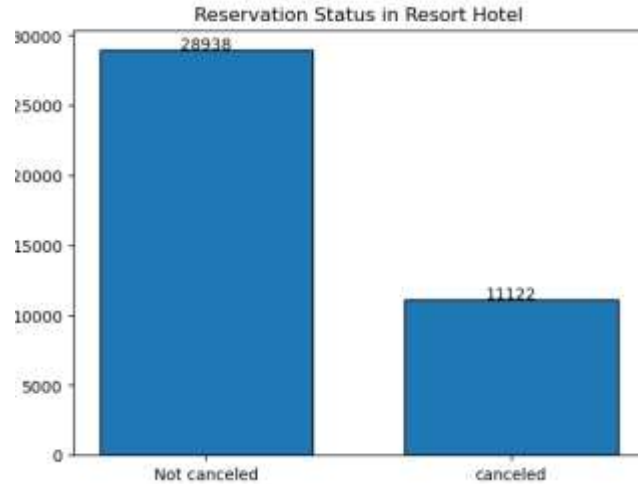
Figure 5. Analysis of hotel bookings – reservation vs. cancellation trends

Business Intelligence Dashboard

The analysis also evaluates the effectiveness of marketing campaigns and loyalty programs. Graphs demonstrate that promotional discounts and special packages significantly increased bookings, particularly for mid-range rooms. Loyalty members showed a higher retention rate and higher average spending, suggesting that tailored loyalty benefits can drive customer loyalty and profitability.
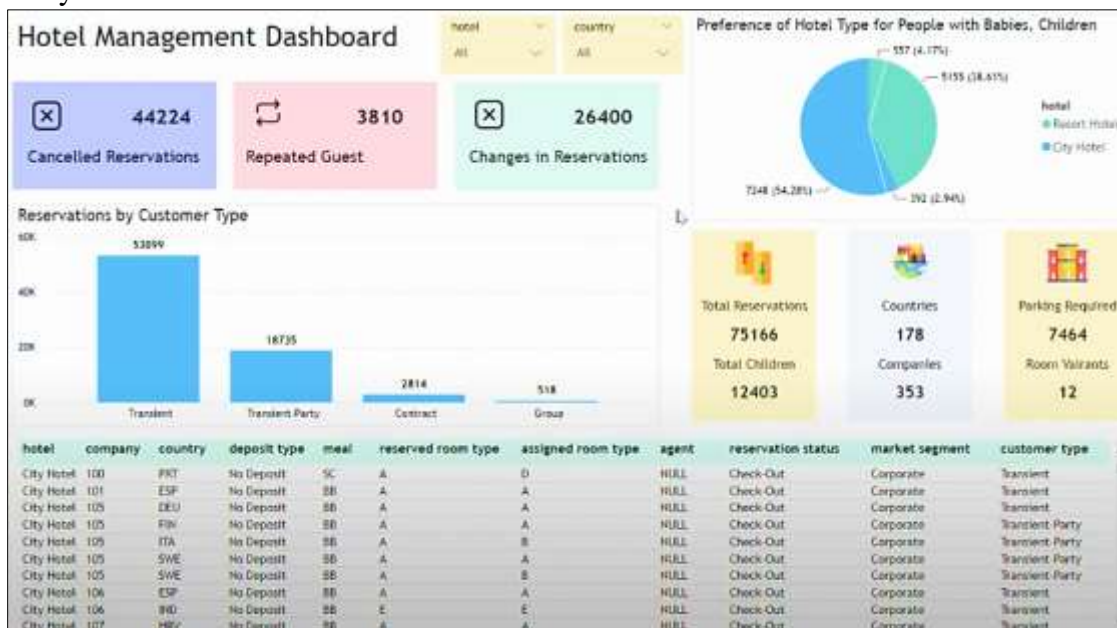

Figure 6. Data visualizaion dashboard for hotel booking management

These experimental results underline the potential of data analytics to transform hotel operations by providing actionable insights into booking behaviors, optimizing marketing efforts, and enhancing customer satisfaction. Figure 6. illustrates interactive data visualization dashboard for managing hotel bookings.

Conclusion

While previous studies such as (Dolnicar et al., 2022) primarily focused on analyzing booking patterns or customer preferences in isolation, our study contributes by developing a comprehensive framework that combines performance metrics, booking history, and customer behavior into a unified data warehouse system. This integration allows for more actionable insights, enabling hotels to optimize their pricing strategies, enhance customer experiences, and improve operational efficiency. Our research thus addresses the gap identified in the literature regarding the lack of a holistic approach to pricing and service delivery in the hospitality industry.

The integration of performance analysis, booking history, and customer behavior offers novel insights into demand forecasting and pricing strategies, which were not fully explored in earlier literature. By developing a unified data-driven system, our research provides a more accurate and actionable framework for hospitality businesses to remain competitive in an increasingly data-driven market.

References

Antonio, N., de Almeida, A., & Nunes, L. (2019). Hotel booking demand datasets. Data in brief, 22, 41-49. DOI: https://doi.org/10.1016/j.dib.2018.11.126

Al-Dmour, R. H., Love, S., & Al-Debei, M. M. (2016). Factors influencing the organisational adoption of human resource information systems: a conceptual model. International Journal of Business Innovation and Research, 11(2), 161-207. DOI: http://dx.doi.org/10.1504/IJBIR.2016.10000012

Badgujar, P. (2023). Enhancing Customer Segmentation Algorithms for Personalized Marketing. Journal of Technological Innovations, 4(4). DOI: https://doi.org/10.93153/j9wfc981

Dolnicar, S. (2022). Market segmentation for e-tourism. In Handbook of e-Tourism (pp. 849-863). Cham: Springer International Publishing. DOI: https://doi.org/10.1007/978-3-030-48652-5_53

Ivanov, M. (2019). The digital marketing with the application of cloud technologies. In SHS Web of Conferences (Vol. 65, p. 04019). EDP Sciences. DOI: https://doi.org/10.1051/shsconf/20196504019

Kimball, R., Ross, M., Becker, B., Mundy, J., & Thornthwaite, W. (2009). Kimball's Data Warehouse Toolkit Classics: The Data Warehouse Toolkit, ; The Data Warehouse Lifecycle, ; The Data Warehouse ETL Toolk. Wiley Publishing.

Sahu, M. K. (2022). Machine Learning for Personalized Marketing and Customer Engagement in Retail: Techniques, Models, and Real-World Applications. Journal of Artificial Intelligence Research and Applications, 2(1), 219-254. Available: https://aimlstudies.co.uk/index.php/jaira/article/view/206

Talluri, K. (2014). New formulations for choice network revenue management. INFORMS Journal on Computing, 26(2), 401-413. DOI: https://doi.org/10.1287/ijoc.2013.0573

Taufik, M., Renaldi, F., & Umbara, F. R. (2021, March). Implementing online analytical processing in hotel customer relationship management. In IOP Conference Series: Materials Science and Engineering (Vol. 1115, No. 1, p. 012040). IOP Publishing. DOI: http://dx.doi.org/10.1088/1757-899X/1115/1/012040

Wardhani, F. Z. D., & Wiratama, J. (2024). Improving the Quality of Service: ETL Implementation on Data Warehouse at Pharmacy Industry. Jurnal Tekno Kompak, 18(1), 1-14. DOI: https://doi.org/10.33365/jtk.v18i1.3211