KOKAND UNIVERSITY HERALD SCIENTIFIC E-JOURNAL

QO‘QON UNIVERSITETI XABARNOMASI
KOKAND UNIVERSITY HERALD
ВЕСТНИК КОКАНДСКОГО УНИВЕРСИТЕТА

ISSN: 2181-1695 **OAK:** 01-08/1819/6 www.herald.kokanduni.uz

ANALYZING HOTEL DATA-DRIVEN SYSTEM BY USING DATA SCIENCE TECHNIQUES

Meliboev Azizjon Ikromjon o‘g‘li,
Faculty of Digital Technologies and Mathematics,
Kokand University

MAQOLA HAQIDA

Qabul qilindi: 24-iyun 2024-yil
Tasdiqlandi: 26-iyun 2024-yil
Jurnal soni: 11
Maqola raqami: 33
DOI: https://doi.org/10.54613/ku.v11i11.971

KALIT SO‘ZLAR/ Ключевые слова/ keywords

Data science, data analysis, Hotel, reservation, cancellation, application, hotel

ANNOTATION

In the past few years, both the City Hotel and Resort Hotel have experienced significant increases in their cancellation rates. As a result, both hotels are currently facing a range of challenges, such as reduced revenue and underutilized hotel rooms. Therefore, the top priority for both hotels is to reduce their cancellation rates, which will enhance their efficiency in generating revenue. This report focuses on the analysis of hotel booking cancellations and other factors that do not directly impact their business and annual revenue generation.

Introduction: In the contemporary hospitality industry, the integration of data science techniques has revolutionized the way hotels operate and make strategic decisions. The vast amounts of data generated by hotel operations, ranging from guest preferences and booking patterns to financial transactions and social media interactions, hold immense potential for driving business insights and enhancing customer experiences. By leveraging data science, hotels can uncover hidden patterns, predict trends, and make data-driven decisions that optimize their operations and boost profitability.

This paper explores the application of data science techniques to analyze hotel data-driven systems. It delves into various methodologies employed in data collection, preprocessing, and analysis to extract valuable insights from hotel data. The focus is on how these insights can be utilized to improve customer satisfaction, streamline operations, and enhance marketing strategies. By examining case studies and real-world applications, the paper highlights the transformative impact of data science in the hospitality sector.

Through a systematic approach, the analysis includes data cleaning, feature engineering, and the application of machine learning models to predict customer behavior, optimize pricing strategies, and personalize marketing campaigns. The integration of advanced analytics and artificial intelligence provides hotels with a competitive edge, enabling them to anticipate market demands, manage resources efficiently, and deliver personalized experiences to their guests.

In summary, the application of data science in hotel data-driven systems represents a significant advancement in the industry, offering a pathway to enhanced operational efficiency and superior customer service. This paper aims to provide a comprehensive understanding of the techniques and benefits of using data science to analyze and improve hotel operations.

Related works: Research on hotel revenue management often involves data science techniques to optimize pricing and maximize revenue. **Talluri and van Ryzin (2004) d**iscusses revenue management in various industries, including hospitality, and highlights the use of forecasting models and optimization techniques to set dynamic pricing. **Bodea and Ferguson (2014) r**eviews revenue management in the hotel industry, emphasizing the importance of demand forecasting and price optimization using historical booking data. **Pang and Lee (2008) pr**ovides a comprehensive overview of sentiment analysis techniques and their applications, including the analysis of online reviews. **Liu, Hu, and Cheng (2017) f**ocuses on sentiment analysis of hotel reviews using machine learning algorithms to gauge customer satisfaction and identify areas for improvement. **Chen, Chiang, and Storey (2012) d**iscusses the role of big data analytics in various industries, including hospitality, and highlights the benefits of predictive analytics for decision-making. **Moreno, García-Alonso, and Fernández-Castro (2016) e**xplores the use of big data techniques for predictive analytics in hotel management, including occupancy and revenue forecasting.

2.1 Dataset description: This dataset contains 119390 observations for a City Hotel and a Resort Hotel. Each observation represents a hotel booking between the 1st of July 2015 and 31st of August 2017, including booking that effectively arrived and booking that were canceled. Features of the data and context of the description is shown in the Table 1.

Table 1.Features of Data

Feature Name	Description
hotel	Type of hotel (Resort hotel or City hotel).
is_canceled	Indicates if the booking was canceled (1) or not (0).
lead_time	Time between the booking date and the arrival date.
arrival_date_year	Year of arrival date.
arrival_date_month	Month of arrival date (12 categories: “January” to “December”).
arrival_date_week_number	Week number of the arrival date.
arrival_date_day_of_month	Day of the month of the arrival date.
stays_in_weekend_nights	Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel.
stays_in_week_nights	Number of week nights (Monday to Friday) the guest stayed.
adults	Number of adults.
children	Number of children.
babies	Number of babies.
meal	Type of meal plan (e.g., BB – Bed & Breakfast).
country	Country of origin.
market_segment	Market segment designation (e.g., “TA” means “Travel Agents” and “TO” means “Tour Operators”).
distribution_channel	Booking distribution channel (e.g., “TA” means “Travel Agents” and “TO” means “Tour Operators”).
is_repeated_guest	Indicates if the booking name was from a repeated guest (1) or not (0).
previous_cancellations	Number of previous bookings that were canceled by the customer prior to the current booking.
previous_bookings_not_can	Number of previous bookings not canceled by the customer prior to the current booking.
reserved_room_type	Code of room type reserved (code is presented instead of designation for anonymity reasons).
assigned_room_type	Code for the type of room assigned to the booking (sometimes differs from the reserved room type due to hotel operation reasons or by customer request, code presented anonymously).
booking_changes	Number of changes/amendments made to the booking.
deposit_type	Type of deposit made (No Deposit, Non Refund, Refundable).
agent	ID of the travel agency that made the booking (ID is presented instead of designation for anonymity reasons).

company	ID of the company/entity that made the booking or responsible for paying the booking (ID is presented instead of designation for anonymity reasons).
days_in_waiting_list	Number of days the booking was in the waiting list before it was confirmed to the customer.
customer_type	Type of customer (Group, Transient, Transient-party).
adr	Average Daily Rate (calculated by dividing the sum of all lodging transactions by the total number of staying nights).
required_car_parking_space	Number of car parking spaces required by the customer.
total_of_special_requests	Number of special requests made by the customer (e.g., twin bed or high floor).
reservation_status	Current reservation status (Check-Out, No-Show).
reservation_status_date	Date at which the last status was set (used in conjunction with reservation_status to understand when the booking was canceled or when the customer checked out).
name	Name of the guest (not real).
email	Email (not real).
phone-number	Phone number (not real).

Methodologies: In the past few years, both the City Hotel and Resort Hotel have experienced significant increases in their cancellation rates. As a result, both hotels are currently facing a range of challenges, such as reduced revenue and underutilized hotel rooms. Therefore, the top priority for both hotels is to reduce their cancellation rates, which will enhance their efficiency in generating revenue. This report focuses on the analysis of hotel booking cancellations and other factors that do not directly impact their business and annual revenue generation. We used Exploratory Data Analysis (EDA) that is a critical techniques in the data analysis process. It involves summarizing the main characteristics of a dataset, often using visual methods. The techniques of EDA can provide to get insights from the dataset, helping to make informed decisions about further data preprocessing, feature selection, and model building. Here is the one of the results that is statistical information of our data in Table 2.

Table 2. Statistical results of Data

[4]:

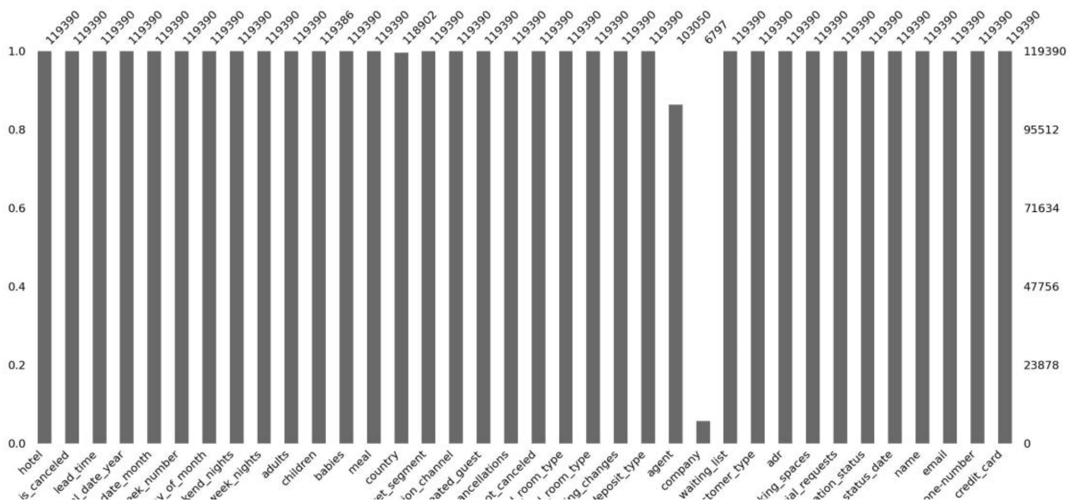
	is_canceled	lead_time	arrival_date_year	arrival_date_week_number	arrival_date_day_of_month	sta
count	119390.000000	119390.000000	119390.000000	119390.000000	119390.000000	
mean	0.370416	104.011416	2016.156554	27.165173	15.798241	
std	0.482918	106.863097	0.707476	13.605138	8.780829	
min	0.000000	0.000000	2015.000000	1.000000	1.000000	
25%	0.000000	18.000000	2016.000000	16.000000	8.000000	
50%	0.000000	69.000000	2016.000000	28.000000	16.000000	
75%	1.000000	160.000000	2017.000000	38.000000	23.000000	
max	1.000000	737.000000	2017.000000	53.000000	31.000000	

Results and Discussion: Data preprocessing is a critical step in the data analysis and machine learning pipeline. It involves transforming raw data into a clean and usable format. This step ensures that the data is consistent, accurate, and suitable for analysis or model training. We used various techniques for analising the data.

- Check Missing values
- Check Duplicates
- Check data type
- Check the number of unique values of each column
- Check statistics of the dataset
- Check various categories present in the different categorical columns

Handling missing values is an essential part of data cleaning during EDA. Here is bar plot of missing values in Figure 1.

Figure 1. Missing values of the data



The provided bar graph illustrates the cancellation and non-cancellation percentages for reservations. It is evident that a substantial portion of reservations remains unaffected by cancellations. Notably, 37% of clients have chosen to cancel their reservations, and this has a noteworthy impact on the hotels' revenue in Figure 2.

Figure 2. Missing values of the data

In comparison to resort hotels, city hotels have more bookings. Its possible that resort hotels are more expensive that those in cities in Figure 3.

Figure 3. Reservation status in different hotels

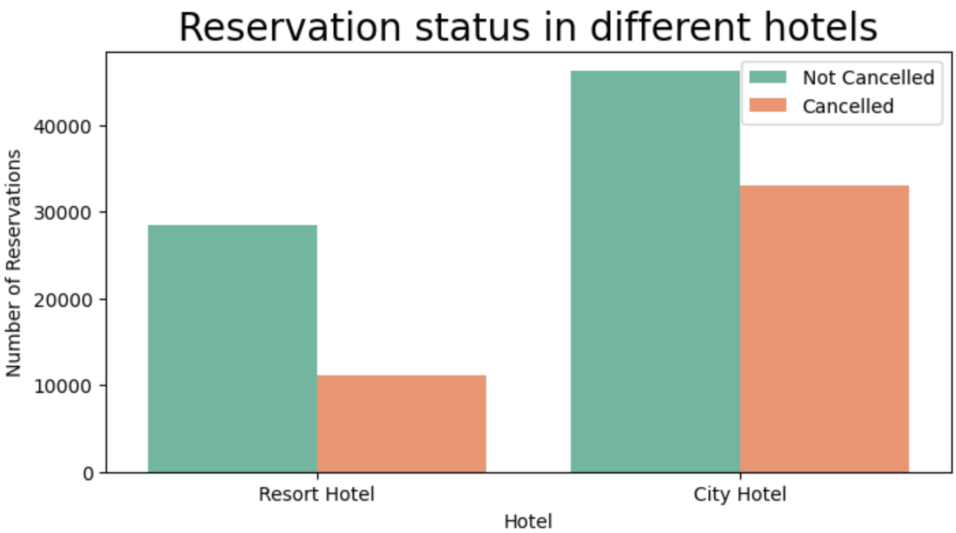
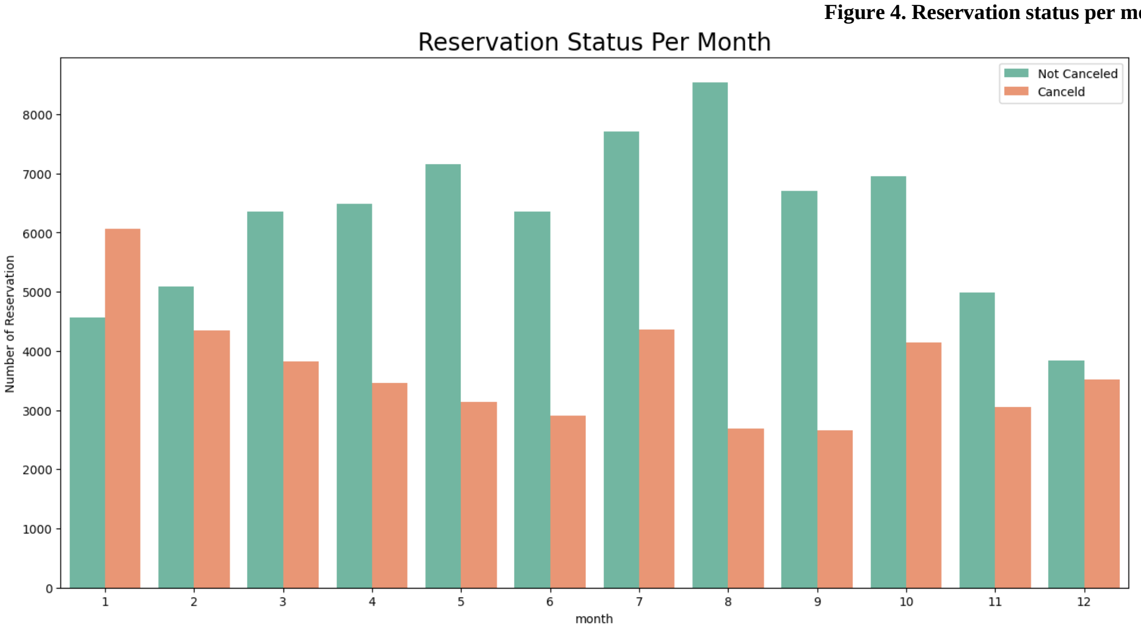


Figure 4. Reservation status per month.



We have created a grouped bar graph to examine the months with the highest and lowest reservation levels based on their status. It's evident that the month of August stands out, having the highest numbers of both confirmed and canceled reservations. In contrast, January has the fewest confirmed reservations but the highest number of canceled reservations in Figure 4.

Conclusion: In this paper, we explored the application of data science techniques to analyze hotel data-driven systems. The hospitality industry generates vast amounts of data from various sources, including booking systems, customer reviews, and operational metrics. By leveraging data science methodologies, hotels can transform this raw data into actionable insights that drive strategic decisions and enhance overall performance. Reservation of the Resort hotels and city hotels is provided by graphs.

References:

1. Meliboev, A., Alikhanov, J., & Kim, W. (2022). Performance evaluation of deep learning based network intrusion detection system across multiple balanced and imbalanced datasets. *Electronics*, *11*(4), 515.
2. Azizjon, M., Jumabek, A., & Kim, W. (2020, February). 1D CNN based network intrusion detection with normalization on imbalanced data. In *2020 international conference on artificial intelligence in information and communication (ICAIIC)* (pp. 218-224). IEEE.
3. Hayotjon o‘g‘li, T. X., & Ikromjon o‘g‘li, M. A. (2023). BIG DATE TIZIMI HAQIDA UMUMIY TASNIF VA TUSHUNCHA. *QO ‘QON UNIVERSITETI XABARNOMASI*, 1281-1284.
4. **Talluri, K.T., & van Ryzin, G.J. (2004).** The Theory and Practice of Revenue Management. Springer Science+Business Media.
5. **Bodea, T., & Ferguson, M. (2014).** Segmentation, Revenue Management and Pricing Analytics. Routledge.
6. **Dolnicar, S., & Otter, T. (2003).** Which hotel attributes matter? A review of previous and a framework for future research. In CAUTHE 2003: Riding the Wave of Tourism and Hospitality Research.
7. **Xia, L., Wang, T., & Zhang, M. (2014).** Examining the influence of recommendation agents on online consumer choice: A personalization-privacy paradox. Decision Support Systems, 57, 93-103.
8. **Ham, S., Kim, W.G., & Jeong, S. (2005).** Effect of information technology on performance in upscale hotels. International Journal of Hospitality Management, 24(2), 281-294.
9. **Sigala, M. (2015).** The application and impact of gamification funware on trip planning and experiences: The case of TripAdvisor’s funware. Electronic Markets, 25(3), 189-209.
10. **Chen, H., Chiang, R.H.L., & Storey, V.C. (2012).** Business intelligence and analytics: From big data to big impact. MIS Quarterly, 36(4), 1165-1188.
11. **Moreno, A., García-Alonso, L., & Fernández-Castro, I. (2016).** Applying predictive analytics in the hospitality industry: A case study of a multinational hotel chain. Journal of Hospitality and Tourism Technology, 7(4), 364-375.