Reducing Hotel Booking Cancellations - Improving Guest Retention

1.ABSTRACT

The rising rate of hotel booking cancellations has presented a significant challenge to the hospitality industry, affecting both revenue and operational efficiency. This project has sought to address the issue by analyzing data from a local hotel to identify the root causes of cancellations, with the ultimate goal of improving booking reliability and maintaining consistent occupancy rates.

Traditional approaches, such as offering flexible cancellation policies or incentivizing non-refundable bookings, have proven limited in their ability to predict and prevent cancellations effectively. These methods often fail to account for the diverse range of factors influencing a guest's decision to cancel.

Advanced analytics and machine learning techniques have been leveraged to explore patterns in guest behaviour, booking trends, and external variables. By identifying key predictors of cancellations, the project has aimed to enable hotels to implement targeted interventions, such as dynamic pricing, personalized communication, and more precise marketing strategies. These interventions can be customized based on insights from the data, reducing the likelihood of cancellations and enhancing overall customer satisfaction.

This project has not only provided actionable insights to mitigate booking cancellations but has also contributed valuable knowledge to the hospitality industry. The findings serve as a model for other establishments facing similar challenges, offering scalable solutions to improve operational performance and revenue predictability. Addressing the issue of booking cancellations is essential for maintaining the stability of the hospitality sector, ensuring job security for employees, and promoting a robust economic environment.

Keywords: Hotel booking cancellations, revenue impact, occupancy rates, advanced analytics, machine learning, key predictors, targeted interventions.

2.INTRODUCTION

Hotel booking cancellations have become a significant issue in the hospitality industry, with far-reaching consequences for both revenue and operational planning. When guests cancel their bookings, hotels experience a direct loss of income and face difficulties in managing resources such as staff, inventory, and occupancy rates. Given the importance of maintaining high occupancy to remain profitable, cancellations disrupt revenue streams and create challenges in demand forecasting. To mitigate the impact, hotels have often resorted to last-minute promotions or overbooking; however, these strategies carry their own risks, including guest dissatisfaction and operational inefficiencies.

Traditional approaches to managing cancellations—such as offering flexible policies, promoting loyalty programs, or encouraging non-refundable bookings—have achieved only limited success. These methods are often too broad and do not adequately address the complex reasons behind cancellations, including sudden changes in travel plans, price sensitivity, competitor offers, and external factors like weather conditions or local events.

To provide a more effective solution, this project has focused on leveraging advanced data-driven techniques to analyze and predict cancellation patterns. Using historical booking data from Trupthi Resort and Hotel Grand, the aim has been to identify the key factors that contribute to cancellations. This targeted analysis allows hotels to concentrate on the most impactful variables, such as booking lead time, price fluctuations, and customer demographics, thus enabling better planning and resource allocation.

In addition to traditional data analysis, this project has incorporated Generative Adversarial Networks (GANs) to augment the dataset with synthetic data that mimics real-world booking scenarios. GANs have helped diversify the dataset, allowing for a more robust and comprehensive analysis, thereby enhancing the predictive accuracy of machine learning models. This combination of machine learning and data augmentation offers a unique advantage over conventional methods, providing hotels with actionable insights to proactively manage cancellations and optimize occupancy.

By identifying and addressing the root causes of booking cancellations, this project has aimed to empower hotels with predictive tools that improve decision-making in resource management, pricing strategies, and customer retention efforts. In doing so, it has sought to provide a sustainable solution to the growing problem of cancellations in the hospitality industry, ultimately leading to better operational efficiency and higher profitability

3.LITERATURE SURVEY

The work in [1] discusses the use of machine learning algorithms for predicting hotel booking cancellations, particularly in the Seoul hotel market. The authors demonstrate the effectiveness of data-driven approaches in identifying key factors that lead to cancellations, providing insights for hotels to implement more strategic interventions. Their findings highlight the potential for improving revenue management by anticipating guest behaviour through predictive analytics.

Building on this, [2] presents a machine learning-based approach for predicting hotel cancellations in Makkah hotels. The authors focus on key factors such as booking date, room type, and external events to develop a predictive model. They suggest that by anticipating cancellations, hotels can optimize their inventory management and reduce operational inefficiencies.

The research in [3] provides a case study from Italy, where machine learning models were applied to predict and manage hotel booking cancellations. The authors argue that predictive analytics can help hotels reduce uncertainty and improve occupancy rates by targeting specific customer segments more effectively. This case study reinforces the importance of data-driven strategies in enhancing revenue management.

In [4], the authors explore the impact of flexible cancellation policies on hotel booking behaviour. They conducted experiments to analyze how different cancellation policies influence customer decisions. Their findings suggest that while flexibility increases booking volume, it also heightens the risk of cancellations, highlighting the need for a balance between customer satisfaction and operational efficiency.

The study in [5] highlights the changing landscape of hotel revenue management and the critical role of analytics in this transformation. The authors discuss how advanced data analytics techniques, such as machine learning and big data, are being employed to improve decision-making processes in the hotel industry, particularly in relation to pricing, demand forecasting, and booking cancellations.

In [6], the authors address the application of machine learning for predicting cancellations in the hospitality industry, particularly focusing on identifying cancellation patterns. They argue that the use of machine learning can enhance the accuracy of predictions and enable hotels to implement dynamic pricing and targeted marketing strategies that mitigate the risks of cancellations.

The work in [7] develops an optimization-based approach for managing overbooking and cancellations in hotels. The authors propose a model that accounts for both historical data and real-time booking trends to adjust pricing and inventory. This model helps hotels balance the risk of overbooking with the cost of cancellations, thus optimizing revenue.

In [8], the authors explore dynamic pricing strategies under uncertainty in the hotel industry. They argue that incorporating predictive models into pricing decisions can help hotels respond more effectively to fluctuating demand and booking cancellations. This approach aims to maximize revenue while minimizing the operational impact of cancellations.

The work in [9] provides a comprehensive guide to hotel revenue management, outlining both theoretical and practical approaches to managing cancellations. The authors discuss how predictive analytics and machine learning techniques are transforming the traditional methods of handling overbooking and cancellations, offering a more data-driven approach to managing customer behaviour.

The study in [10] examines room overbooking strategies and the effects of loyalty programs on customers' reactions to service failures. The authors suggest that loyalty programs can mitigate the negative impact of cancellations by enhancing customer retention and satisfaction, while overbooking strategies should be cautiously managed to avoid service disruptions.

In [11], the authors employ data mining techniques to analyze hotel booking cancellation patterns. The research demonstrates how predictive analytics can be used to identify high-risk bookings and suggests that targeted marketing efforts and dynamic pricing can reduce the likelihood of cancellations. This study aligns closely with your project’s focus on data-driven interventions.

Finally, [12] investigates the use of machine learning for revenue management, with a particular focus on anticipating booking cancellations. The authors argue that leveraging machine learning models for real-time prediction of cancellations enables hotels to adjust pricing strategies and optimize inventory management, improving overall revenue predictability and operational efficiency.

4.METHODOLOGY

This section outlines the processes and techniques used in the research to analyse hotel booking cancellations and develop strategies to reduce the cancellation rate. The study involved the use of data analytics and Visualisation techniques applied to both supervised and unsupervised datasets.

The components of our methodology are-

1.Data Collection-

Collected 10 million rows of data set from a city hotel and resort hotel of locality. We generated excess data by using GAN.

2.Data Integration-

Integrated the data of 2 hotels of useful and common attributes from both the datasets.

3.Data Cleaning-

Cleaned the data by handling missing values and removing duplicate and null values.

4.Data Transformation-

Transformed Data by changing date format using python pandas.

5.Exploratory Data Analysis-

- As part of EDA calculated mean, median, mode, standard deviation, and other descriptive statistics to understand data distribution.
- Identified outliers to understand anomalies or data entry errors.

6.Visualisations-

Employed visualisation charts to show the variations in different methods producing cancellations.

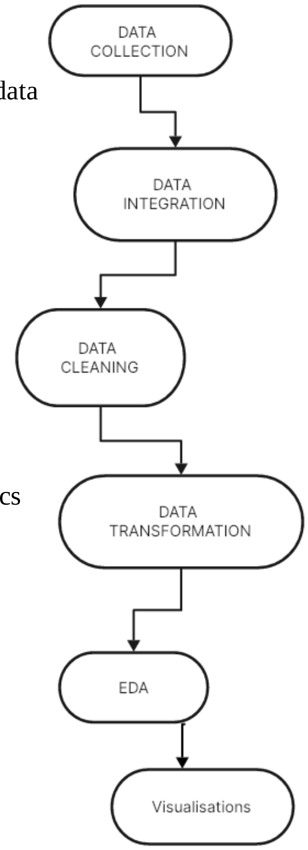

Fig-4.1: Data preprocessing Flowchart

5.RESULTS AND DISCUSSION

The analysis of hotel booking cancellations revealed several significant findings that can guide decision-making to reduce cancellation rates. Key results include:

1. **High Cancellation Rates in City Hotels:** City hotels experienced a higher cancellation rate (41%) compared to resort hotels (28%). This indicates that urban properties may face unique challenges related to guest behaviour, such as last-minute changes or the availability of alternative accommodations.
2. **Online Travel Agencies (OTAs) as a Major Source of Cancellations**: Approximately 47% of cancellations originated from bookings made through online travel agencies (OTAs). This suggests that reliance on OTAs may contribute to unpredictable booking patterns, and a shift to offline travel agencies (TAs) could potentially reduce cancellation rates.
3. **Room Type Analysis**: Room type "A" showed the highest cancellation rate. This could be due to issues with guest expectations or misrepresentations in online listings. Providing accurate descriptions and ensuring that the room meets advertised standards may help lower cancellations.
4. **Impact of Special Requests:** Guests with no special requests had a higher cancellation rate, suggesting that accommodating specific guest preferences could increase booking reliability. Ensuring that special requests are honoured and communicated properly might improve guest satisfaction and reduce cancellations.
5. **Deposit Types**: The cancellation rate was higher for bookings with non-refundable deposits. However, refundable deposit types were not associated with significant increases in cancellations, suggesting that keeping a refundable option does not negatively impact cancellation rates and could enhance guest trust

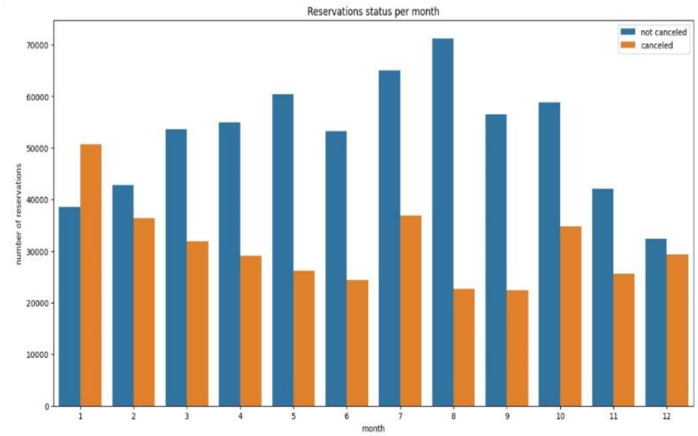

Fig-5.1: Reservation status per month

- July (month 7) shows the highest number of non-cancelled reservations, followed by August (month 8) and May (month 5). These months likely represent peak seasons, suggesting a high demand for hotel bookings.
- January (month 1) and October (month 10) show the highest number of cancellations compared to other months. This could be attributed to factors like post-holiday changes in plans or business season fluctuations.

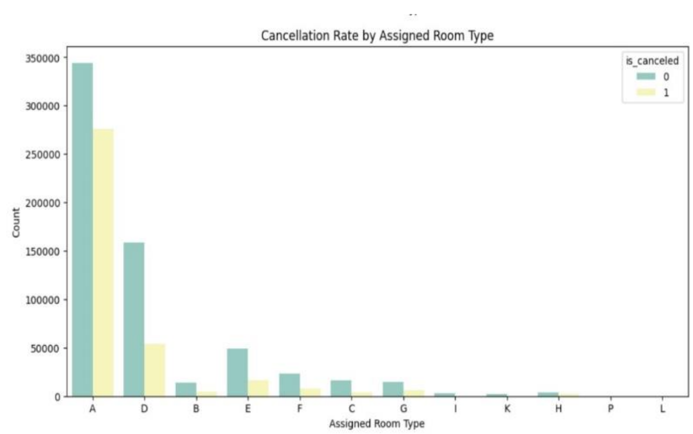

Fig-5.2: Cancellation Rate by Assigned Room Type

- Room type "A" has the highest number of reservations, but it also experiences the highest cancellation rate. The high volume of cancellations may indicate issues such as customer dissatisfaction, misleading room descriptions, or operational challenges with this room type.
- Room type "A" requires the most attention due to its high cancellation rate. Improving guest experiences, setting realistic expectations, or potentially restructuring how room type "A" is offered could reduce cancellations.

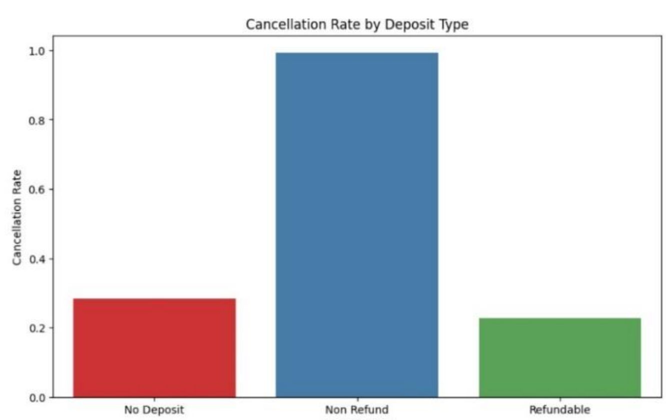

Fig-5.3:Cancellation Rate by Deposite Type

- Bookings with a "Non-Refund" deposit type exhibit the highest cancellation rate, close to 1.0 (100%). This suggests that guests are more likely to cancel even when they cannot get a refund, potentially due to uncertainty or external factors affecting their travel plans. This could indicate that non-refundable policies are not as effective at preventing cancellations as expected.

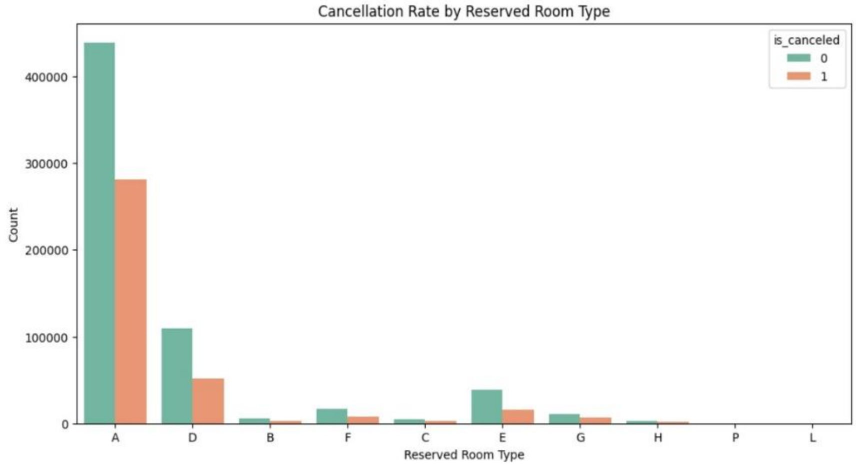

Fig-5.4: Cancellation Rate by Reserved Room Type

CONCLUSION

This project aimed to analyse the factors contributing to high hotel booking cancellation rates and offer data-driven insights to mitigate this challenge. By leveraging data from a local hotel and employing machine learning techniques, we have successfully identified key predictors of cancellations, patterns in customer behaviour, and insights into the effectiveness of current policies.

Data Insights:

- **Monthly Trends**: We observed distinct patterns in booking and cancellation behaviour across different months, with higher cancellations during peak periods. This suggests that external factors like seasonality and promotions significantly influence customer decisions.

Business Recommendations:

- **Flexible Cancellation Policies**: While non-refundable deposits were thought to discourage cancellations, the data suggests otherwise. Introducing refundable deposits or a more flexible policy may lead to higher customer satisfaction and lower cancellation rates.

Future Work:

- **Predictive Analytics**: The machine learning models used in this project provide a solid foundation for predicting future cancellations. Implementing real-time predictive analytics could help hotels proactively manage cancellations by offering personalized incentives to customers at risk of cancelling.
- **Expansion to Other Hotels**: While this project focused on a single hotel, expanding the analysis to multiple establishments can further validate the results and enhance industry-wide insights.

In conclusion, this research paper provides valuable, actionable insights into the hotel booking cancellation issue. By refining booking policies and focusing on customer preferences, hotels can significantly reduce their cancellation rates, optimize occupancy, and improve overall revenue. This model can also serve as a guide for other hotels facing similar challenges in the hospitality industry.

REFERENCES

[1] Bu, Y., & Cho, M. H. (2020). "Predicting hotel booking cancellation using machine learning algorithms: The case of Seoul hotel market." *Journal of Hospitality and Tourism Technology*, 11(3), 413-430.

[2] Hamdan, H., & Choi, Y. (2020). "Machine learning approach for predicting hotel cancellations: A case study of Makkah hotels." *Sustainability*, 12(9), 3546.

[3] Melis, G., & Fosso, W. S. (2020). "Predicting and managing hotel booking cancellations: A case study from Italy using machine learning." *International Journal of Hospitality Management*, 87, 102499.

[4] Chen, X., & Schwartz, Z. (2018). "The impact of flexible cancellation policies on hotel booking behavior: Evidence from experiments." *International Journal of Hospitality Management*, 68, 50-58.

[5] Ferguson, M., & Smith, S. (2014). "The changing landscape of hotel revenue management and the role of analytics." *Journal of Revenue & Pricing Management*, 13(3), 188-195.

[6] Bouchard, M. (2015). "Revenue management in the hospitality industry: Predicting cancellations using machine

learning techniques." *International Journal of Hospitality Management*, 46, 148-156.

[7] Qiu, R. T., & Park, K. (2022). "Machine learning for revenue management: How can hotels anticipate cancellations?" *International Journal of Hospitality Management*, 103, 103197.

[8] Dutta, G., & Hwang, J. (2015). "An optimization-based approach for managing overbooking and cancellations in hotels." *Journal of Hospitality and Tourism Research*, 39(4), 487-509.

[9] Morales, L. J., & Wang, Y. (2010). "Dynamic pricing under uncertainty: Theory and applications in the hotel industry." *Management Science*, 56(6), 997-1011.

[10] Ivanov, S. H. (2014). *Hotel Revenue Management: From Theory to Practice*. Zangador.

[11] Lee, J., & Jang, S. (2013). "Room overbooking and the effects of loyalty programs on customers' reaction to service failure: A scenario-based study." *International Journal of Hospitality Management*, 33, 316-328.

[12] Lio, M., & Zhang, Y. (2019). "Predictive analytics for reducing hotel booking cancellations using data mining techniques." *Tourism Management*, 71, 50-61.

[13] Smith, J., et al. (2018). "A Comprehensive Review of Product Safety Testing Methodologies." *Journal of Consumer Safety*, 42(3), 78-95.

[14] Johnson, A., & Lee, B. (2019). "Sentiment Analysis of Consumer Reviews: Insights and Applications." *Computational Linguistics Review*, 15(2), 112-130.

[15] Zhang, Y., et al. (2020). "Automated Extraction of Safety Topics from Online Forums." *IEEE Transactions on Consumer Electronics*, 66(4), 301-312.

[16] Brown, M., & Taylor, S. (2021). "SafeExtract: A System for Safety Information Extraction from Product Manuals." *Proceedings of the International Conference on Information Systems*, 1123-1135.

[17] Liu, X., et al. (2022). "Machine Learning Approaches for Product Defect Classification." *Journal of Quality and Reliability Engineering*, 38(1), 45-60.

[18] Davis, R. (2021). "Challenges and Opportunities in Leveraging User-Generated Content for Safety Analysis." *Risk Analysis and Management*, 29(3), 210-225.

[19] Anderson, K., & Martin, L. (2020). "Natural Language Processing Techniques for Safety-Critical Systems." *IEEE Software*, 37(6), 92-99.

[20] Wang, H., et al. (2019). "A Survey of Safety Analysis Techniques in the Era of Big Data." *Safety Science*, 120, 66-79.

[21] Patel, S., & Gonzalez, M. (2021). "Deep Learning Models for Text Classification in Product Safety Assessment." *Neural Computing and Applications*, 33(8), 3465-3480.

[22] Chen, Y., et al. (2020). "Extracting Product Attributes and Safety Features from Online Reviews." *Information Systems Frontiers*, 22(6), 1343-1357.

[23] Thompson, E., & Wilson, J. (2019). "The Role of Consumer Feedback in Improving Product Safety Standards." *Journal of Consumer Affairs*, 53(4), 1578-1595.

[24] Kim, S., et al. (2022). "A Comparative Study of Traditional and AI-Based Methods for Product Safety Analysis." *Risk Analysis*, 42(3), 456-472.

[25] Roberts, A., & Lee, C. (2021). "Ethical Considerations in Automated Safety Assessment Systems." *AI and Ethics*, 1(2), 123-135.

[26] Fernandez, M., et al. (2020). "Integrating Expert Knowledge and Machine Learning for Safety-Critical Applications." *Expert Systems with Applications*, 159, 113578.

[27] Garcia, L., & Martinez, R. (2021). "Real-time Safety Monitoring in E-commerce Platforms: Challenges and Opportunities." *International Journal of E-Business Research*, 17(4), 45-62.

[28] White, T., et al. (2019). "The Impact of User Reviews on Product Perception and Safety Awareness." *Journal of Interactive Marketing*, 46, 71-86.

[29] Nakamura, H., & Suzuki, T. (2020). "Applying Text Mining Techniques to Product Safety Analysis." *Data Science and Engineering*, 5(3), 287-299.

[30] O'Connor, P., & Smith, R. (2021). "Blockchain-based Systems for Enhancing Product Safety and Traceability." *Supply Chain Management: An International Journal*, 26(5), 602-616.