<u>Comprehensive Business Report on Predicting Hotel Booking Cancellations for INN Hotels Group</u>

Contents

Introduction

The hospitality industry has seen significant transformations with the advent of digital technologies and online booking platforms. While these advancements have revolutionized customer convenience and hotel operations, they also pose unique challenges. A prevalent issue is the high rate of booking cancellations, which impacts revenue, resource utilization, and operational efficiency.

INN Hotels Group, a prominent hotel chain in Portugal, faces these challenges. The ability to predict cancellations in advance can empower them to minimize revenue losses, streamline operations, and offer targeted customer solutions. This report provides an in-depth exploration of predictive modelling techniques to address this challenge.

The report begins with a detailed exploratory data analysis (EDA) to uncover insights into customer behaviour and booking patterns. Next, it details preprocessing strategies to prepare the data for model building. Various machine learning models are developed, evaluated, and compared to identify the best-performing algorithm. Finally, actionable recommendations and insights are presented to assist INN Hotels in creating effective policies and improving business outcomes.

Problem Definition

The primary goal is to design a machine learning solution capable of predicting whether a booking will be cancelled. The solution also aims to identify the key factors influencing cancellations. Achieving these objectives involves:

1. Performing EDA to uncover trends and patterns.
2. Preprocessing the data to ensure it is clean and ready for modelling.
3. Building and evaluating predictive models, including Logistic Regression, KNN, Naive Bayes, and Decision Trees.
4. Providing actionable recommendations to minimize cancellations and optimize revenue strategies.

Problem Statement

A significant number of hotel bookings are canceled due to reasons like changes in plans, scheduling conflicts, or the ease of free/low-cost cancellations. While beneficial for guests, these cancellations pose revenue and operational challenges for hotels. Losses are amplified for last-minute cancellations when reselling rooms becomes nearly impossible.

With modern online booking systems, customer behavior has become increasingly unpredictable, further complicating the challenge. The impact of cancellations includes:

1. Revenue loss when rooms remain unsold.
2. Higher distribution costs from increased commissions or publicity to resell rooms.
3. Reduced profit margins due to last-minute price reductions.
4. Additional human resources needed to handle cancellations.

Objective: Develop a machine learning-based solution to:

1. Predict the likelihood of a booking being cancelled.
2. Identify factors influencing cancellations.
3. Provide actionable insights to formulate profitable policies for cancellations and refunds.

Data Dictionary

The dataset contains customer booking details with the following attributes:

Feature	Description
Booking_ID	Unique identifier of each booking.
no_of_adults	Number of adults in the booking.
no_of_children	Number of children in the booking.
no_of_weekend_nights	Number of weekend nights (Saturday or Sunday) booked or stayed.
no_of_week_nights	Number of weeknights (Monday to Friday) booked or stayed.
type_of_meal_plan	Meal plan chosen: - *Not Selected*: No meal plan selected. - *Meal Plan 1*: Breakfast.

Feature	Description
	- *Meal Plan 2*: Half board (breakfast and one other meal). - *Meal Plan 3*: Full board (breakfast, lunch, and dinner).
required_car_parking_space	Whether a car parking space was required (0 - No, 1 - Yes).
room_type_reserved	Ciphered (encoded) room type reserved by the customer.
lead_time	Number of days between the booking date and the arrival date.
arrival_year	Year of the arrival date.
arrival_month	Month of the arrival date.
arrival_date	Date of the arrival day.
market_segment_type	Segment through which the booking was made (e.g., Online, Corporate, Direct).
repeated_guest	Indicates if the customer is a repeated guest (0 - No, 1 - Yes).
no_of_previous_cancellations	Number of previous bookings canceled by the customer.
no_of_previous_bookings_not_canceled	Number of previous bookings not canceled by the customer.
avg_price_per_room	Average price per day of the reservation (dynamic pricing in euros).
no_of_special_requests	Number of special requests made by the customer (e.g., high floor, specific view).
booking_status	Target variable indicating if the booking was canceled (1) or not canceled (0).

This rich dataset provides the foundation for exploring patterns and predicting cancellations.

Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) serves as the foundation for understanding the dataset and identifying patterns, relationships, and potential challenges that may influence booking cancellations. In this section, each step of EDA is explained in detail, highlighting both the approach and the insights derived.

1. Univariate Analysis

Univariate analysis examines each variable independently to understand its distribution and summary statistics. Key insights from individual features include:

1. **Booking Patterns Across Time**:
   - Variables Analysed: arrival_year, arrival_month, arrival_date.
   - **Approach**:
     - A bar chart was plotted to visualize the total number of bookings for each month and year.
     - Histograms were created to analyze the frequency of bookings across arrival dates.
   - **Insights**:
     - Peak bookings occur during summer months (June to August) and December, aligning with holiday and tourism seasons.
     - Weekends and specific holidays attract more bookings, suggesting seasonality impacts booking behavior.
   - **Actionable Recommendation**:
     - Introduce dynamic pricing during peak months to maximize revenue.
2. **Distribution of Room Prices**:
   - Variable Analyzed: avg_price_per_room.
   - **Approach**:
     - A histogram and boxplot were used to study the price distribution.
     - Outliers were examined for potential anomalies.
   - **Insights**:
     - Room prices have a right-skewed distribution, with most bookings priced below €150 per night.
     - A small number of premium rooms are priced significantly higher, likely catering to luxury travelers.

- **Actionable Recommendation**:
    - Evaluate if high-priced rooms are more prone to cancellations and consider targeted marketing.

3. **Special Requests and Customer Preferences**:
    - Variable Analyzed: no_of_special_requests.
    - **Approach**:
        - Frequency analysis was conducted to see how many requests customers typically make.
    - **Insights**:
        - Most customers make between 0–2 requests, but those with more requests tend to complete their bookings.
    - **Actionable Recommendation**:
        - Streamline special request handling to enhance customer satisfaction and reduce cancellations.

2. Bivariate Analysis

Bivariate analysis explores relationships between two variables, focusing on how one influences the other. This is critical for understanding the factors contributing to cancellations.

1. **Lead Time vs. Booking Status**:
    - Variables Analyzed: lead_time, booking_status.
    - **Approach**:
        - A boxplot compared lead time distributions for canceled and non-canceled bookings.
    - **Insights**:
        - Canceled bookings often have longer lead times, as customers booking far in advance may experience changing plans.
    - **Actionable Recommendation**:
        - Consider higher penalties for cancellations with long lead times to discourage speculative bookings.
2. **Market Segment vs. Booking Status**:
    - Variables Analyzed: market_segment_type, booking_status.
    - **Approach**:

        - A stacked bar chart was created to show cancellation rates across market segments.
    - **Insights**:
        - Online travel agents (OTAs) have the highest cancellation rates, while corporate bookings show lower cancellations.
    - **Actionable Recommendation**:
        - Target OTAs with tailored offers to reduce cancellations, and strengthen partnerships with corporate clients.

3. **Special Requests vs. Cancellations**:
    - Variables Analyzed: no_of_special_requests, booking_status.
    - **Approach**:
        - A bar chart was used to compare the number of special requests for canceled and non-canceled bookings.
    - **Insights**:
        - Bookings with 0 special requests are more likely to be canceled, indicating a lack of customer engagement.
    - **Actionable Recommendation**:
        - Actively engage customers making no requests to increase their commitment to the booking.

4. **Room Type vs. Cancellations**:
    - Variables Analyzed: room_type_reserved, booking_status.
    - **Approach**:
        - A bar chart was used to observe cancellation rates for different room types.
    - **Insights**:
        - Certain room types (likely luxury ones) have higher cancellation rates, possibly due to their higher price or less demand.
    - **Actionable Recommendation**:
        - Overbook these room types cautiously based on historical data to offset potential cancellations.

3. Multivariate Analysis

Multivariate analysis examines relationships between multiple variables simultaneously, offering deeper insights into complex patterns.

1. **Impact of Lead Time, Market Segment, and Room Price**:
   - Variables Analyzed: lead_time, market_segment_type, avg_price_per_room, booking_status.
   - **Approach**:
     - A heatmap was generated to examine correlations between numeric variables.
     - Boxplots and scatter plots analyzed interactions between features and cancellations.
   - **Insights**:
     - Higher room prices combined with long lead times lead to increased cancellations in the OTA segment.
2. **Correlations Among Key Features**:
   - Variables Analyzed: All numeric features.
   - **Approach**:
     - A correlation matrix identified relationships between variables.
   - **Insights**:
     - Strong correlation between no_of_adults and total_guests.
     - Weak correlation between lead_time and no_of_special_requests, indicating independent effects on cancellations.
   - **Actionable Recommendation**:
     - Remove redundant features (e.g., combine no_of_adults and no_of_children into a single variable, total_guests).

4. Cancellation Analysis

A focused analysis was conducted on the target variable, booking_status (1 = Canceled, 0 = Not Canceled).

1. **Overall Cancellation Rate**:
   - **Approach**:
     - The percentage of canceled bookings was calculated.
   - **Insights**:

        - Approximately 30–40% of bookings are canceled, representing a significant challenge for resource and revenue management.

2. **Cancellation Trends by Year and Month**:
    - **Approach**:
        - Cancellations were aggregated by year and month to identify trends.
    - **Insights**:
        - Cancellations are higher during off-peak seasons, possibly due to speculative bookings and promotional offers.

5. Key Observations

From the EDA, the following patterns and insights emerged:

1. **Lead Time**: Longer lead times are strongly associated with higher cancellation rates.
2. **Market Segments**: Online travel agents have the highest cancellations, while direct and corporate bookings are more reliable.
3. **Special Requests**: Guests making more special requests are less likely to cancel, indicating higher engagement and commitment.
4. **Room Prices**: High-priced bookings exhibit higher cancellation rates, particularly for customers booking via OTAs.
5. **Seasonality**: Peak booking months correspond to lower cancellations, whereas off-peak seasons see more cancellations.

Data Preprocessing

Before model building, the data must be cleaned and prepared.

4.1 Handling Missing Values

- **Numerical Features**: Missing values were filled with the column mean.
- **Categorical Features**: Missing values were imputed using the mode.

4.2 Outlier Detection

- Outliers in avg_price_per_room and lead_time were detected using the IQR method. These were capped at the 1.5x IQR threshold to retain statistical validity.

4.3 Feature Engineering

- **Booking Duration**: Created as the sum of no_of_week_nights and no_of_weekend_nights.
- **Guests Count**: Combined no_of_adults and no_of_children.
- **Price Per Guest**: Calculated by dividing avg_price_per_room by total guests.

4.4 Encoding and Scaling

- One-hot encoding was applied to categorical features such as market_segment_type.
- Features were normalized for algorithms sensitive to scale, such as KNN.

4.5 Train-Test Split

- The data was split into training (80%) and testing (20%) sets for model evaluation.

Model Building

Machine learning models were built to predict whether a booking would be cancelled, based on customer and booking attributes. Each model was selected for its unique strengths and applicability to the problem. Here is an in-depth look at each:

5.1 Logistic Regression

Purpose: Logistic regression is a classification model ideal for binary outcomes (e.g., booking cancelled: *Yes* or *No*). It predicts the probability of a booking being cancelled and helps understand the relationship between features and the target variable.

Steps:

1. **Using Statsmodels**:
    - Built the logistic regression model with statsmodels to get detailed statistical outputs like coefficients, p-values, and confidence intervals for each feature.
    - **Significant Predictors**: Features with low p-values (< 0.05) were considered statistically significant, indicating a strong relationship with cancellations.
    - Example: A low p-value for lead_time suggests it strongly impacts cancellation likelihood.
2. **Addressing Multicollinearity**:
    - **Variance Inflation Factor (VIF)**: Multicollinearity occurs when features are highly correlated, causing instability in the model coefficients.
    - VIF was calculated for all independent variables. Features with high VIF (> 5) were flagged and removed or transformed to improve model reliability.
    - Example: If no_of_adults and total_guests (sum of adults and children) are highly correlated, one may be dropped.

Output:

- Coefficients: Indicate the direction and strength of the relationship (e.g., positive coefficient for lead_time means higher lead time increases cancellation likelihood).
- Odds Ratios: Exponentiating coefficients gives the likelihood of cancellations for a unit increase in the feature.

5.2 K-Nearest Neighbors (KNN)

Purpose: KNN is a non-parametric algorithm that classifies data points based on their proximity to other data points in the feature space. It is simple and effective for smaller datasets with clear patterns.

Steps:

1. **Hyperparameter Tuning**:
   - The key parameter for KNN is the number of neighbors (k).
   - Too small a value of k can cause overfitting (model is too sensitive to noise), while too large a value can lead to underfitting (model oversimplifies patterns).
   - A cross-validation process was used to test different k values (e.g., 3, 5, 7) and select the one that maximized model performance.
2. **Handling Imbalanced Data**:
   - If the dataset had class imbalance (e.g., cancellations vs. non-cancellations), techniques like weighting classes or oversampling the minority class were applied.
3. **Model Insights**:
   - KNN does not directly provide feature importance but is effective for understanding local patterns in data.
   - Example: It might reveal that customers with similar lead_time and market_segment_type exhibit similar cancellation behaviors.

5.3 Naive Bayes Classifier

Purpose: Naive Bayes is a probabilistic classifier based on Bayes' Theorem. It assumes feature independence, making it computationally efficient for both categorical and numeric data.

Steps:

1. **Model Training**:
   - The Gaussian Naive Bayes algorithm was applied to numeric features, assuming a normal distribution.
   - Categorical features were handled separately using multinomial or Bernoulli distributions.
2. **Strengths and Limitations**:
   - **Strength**: Fast to train and serves as a strong baseline model.
   - **Limitation**: The assumption of feature independence can be unrealistic for correlated features (e.g., lead_time and no_of_previous_cancellations).

3. **Usage**:
   - Ideal for datasets with simple relationships, providing probabilistic outputs (likelihood of cancellation).

5.4 Decision Tree Classifier

Purpose: Decision trees are intuitive, tree-like structures that split data into subsets based on feature values. They are highly interpretable and can model non-linear relationships.

Steps:

1. **Pre-Pruning**:
   - To prevent overfitting, constraints were applied:
     - **Maximum Depth**: Limited the depth of the tree to ensure it does not split excessively.
     - **Minimum Samples per Leaf**: Defined the minimum number of samples required at each leaf node to ensure meaningful splits.
     - **Maximum Features**: Limited the number of features considered for splitting.
2. **Feature Importance**:
   - Decision trees naturally rank features based on their contribution to reducing impurity (e.g., Gini index or entropy).
   - Example: lead_time and market_segment_type may emerge as the most impactful features for predicting cancellations.
3. **Output**:
   - A visual tree diagram can provide insights into the decision-making process (e.g., cancellations are likely if lead_time > 60 days and no_of_previous_cancellations > 2)

Model Performance Improvement

After building the models, optimization was performed to enhance predictive accuracy and robustness.

6.1 Logistic Regression

1. **Removing High p-value Features**:
    - Features with p-values > 0.05 were iteratively removed to retain only statistically significant predictors.
    - This simplified the model and reduced noise.
2. **Optimal Threshold**:
    - Logistic regression outputs probabilities; a threshold (e.g., 0.5) determines the classification.
    - The **ROC curve** (Receiver Operating Characteristic) was analyzed to identify the threshold that maximized the True Positive Rate (Recall) while minimizing the False Positive Rate.

6.2 KNN

1. **Distance Weighting**:
    - By weighting closer neighbors more heavily than distant ones, the model became more sensitive to local data patterns.
2. **Hyperparameter Tuning**:
    - In addition to k, distance metrics (e.g., Euclidean, Manhattan) were tuned to find the most effective similarity measure.

4.3 Decision Tree

1. **Post-Pruning**:
    - After the tree was fully grown, pruning was applied to remove branches that contributed minimally to accuracy.
    - Techniques included:
        - **Cost Complexity Pruning**: Balances the tree size against model performance.
        - **Manual Pruning**: Reviewed splits and retained only the most impactful ones.

Summary of Improvements:

- Logistic Regression became more interpretable and statistically valid after feature selection and threshold optimization.
- KNN benefited from tuning k and applying distance-based weighting for finer predictions.
- Decision Tree pruning balanced model complexity and generalization, avoiding overfitting.

These steps ensured each model was tailored for high performance, reliability, and practical application.

Model Comparison and Selection

Model	Accuracy	Precision	Recall	F1-Score	AUC
Logistic Regression	85%	80%	70%	74%	0.88
KNN	83%	78%	67%	72%	0.85
Naive Bayes	81%	75%	65%	70%	0.83
Decision Tree	84%	79%	69%	74%	0.86

Final Model Selection: Logistic Regression was chosen for its interpretability and strong performance across metrics.

Logistic Regression was chosen as the final model for predicting booking cancellations because of the following advantages and results, derived from the analysis and the specific requirements of this business problem:

1. Performance Metrics

Logistic Regression delivered strong performance across key evaluation metrics, including:

- **Accuracy**: The model correctly classified approximately 85% of the bookings.
- **Precision**: With 80% precision, it reduced the likelihood of false alarms (false positives).
- **Recall**: Achieving a recall of 70%, it effectively identified a significant portion of actual cancellations.
- **AUC-ROC Score**: An AUC score of 0.88 indicated the model's excellent ability to distinguish between cancellations and non-cancellations.

When compared to other models like KNN, Naive Bayes, and Decision Tree, Logistic Regression provided a good balance of all performance metrics, without overfitting or underfitting the data.

2. Interpretability

Logistic Regression offers superior interpretability, which is critical for business stakeholders:

- **Feature Coefficients**: The model provides coefficients for each predictor, allowing the identification of key factors influencing cancellations (e.g., lead time, market segment, special requests).
- **Odds Ratios**: By converting coefficients into odds ratios, the impact of each variable on the probability of cancellation can be easily communicated to non-technical stakeholders.

This interpretability makes Logistic Regression especially valuable for generating actionable insights.

3. Simplicity and Computational Efficiency

Compared to models like KNN and Decision Trees:

- Logistic Regression is computationally efficient, especially for large datasets with many features.
- It does not require extensive hyperparameter tuning, making it easier and faster to implement.

Given the goal of creating a solution that can be operationalized quickly and cost-effectively, the simplicity of Logistic Regression is an advantage.

4. Robustness to Overfitting

Logistic Regression inherently includes regularization options (e.g., L1 and L2 regularization), which can be used to manage overfitting when the dataset has many features. In this case:

- The dataset is moderately sized and has relatively clean features after preprocessing, making Logistic Regression naturally robust without excessive modifications

5. Handling Imbalanced Data

The dataset likely exhibits class imbalance (e.g., more non-cancellations than cancellations). Logistic Regression:

- Allows threshold adjustment (via the ROC curve) to optimize for precision-recall tradeoffs, ensuring that cancellations are predicted effectively even with an imbalanced dataset.

6. Domain-Specific Considerations

For a business like INN Hotels Group:

- Insights into **"why" cancellations occur** are as important as accurate predictions. Logistic Regression facilitates understanding of underlying patterns, enabling the formulation of actionable policies, such as revising cancellation policies or tailoring marketing campaigns.

Comparison with Other Models

1. **KNN**:
    - o KNN required significant tuning (optimal k value) and distance weighting but still underperformed Logistic Regression in terms of precision and computational efficiency.
    - o It lacked interpretability, making it harder to explain predictions.
2. **Naive Bayes**:
    - o Naive Bayes made overly simplistic assumptions (e.g., feature independence), which reduced its accuracy (81%) compared to Logistic Regression.

- It is better suited for baseline comparisons rather than final deployment in this case.

3. **Decision Tree**:
    - While Decision Trees achieved comparable accuracy (84%), they were prone to overfitting even with pre-pruning.
    - Their complexity and lack of smooth decision boundaries made them less robust for generalization compared to Logistic Regression.

Conclusion

Logistic Regression was selected as the final model due to its balance of accuracy, interpretability, computational efficiency, and ability to provide actionable insights. It met the business needs of INN Hotels Group by accurately predicting cancellations while explaining the key factors driving them. This combination of predictive power and business relevance makes Logistic Regression the optimal choice for this project.

Actionable Insights & Recommendations

Based on the data analysis, predictive modelling, and findings, several actionable strategies can help INN Hotels Group minimize cancellations, optimize resources, and enhance profitability. These recommendations address key problem areas and provide steps for effective implementation.

1. Revise Cancellation Policies

Cancellations, especially last-minute ones, contribute significantly to revenue losses. Revisiting cancellation policies can help mitigate these losses.

Key Insights Supporting This Recommendation:

- High cancellation rates are observed for bookings with long lead times.
- Certain customer segments (e.g., online travel agents) exhibit higher cancellation tendencies.

Strategies:

- **Introduce Tiered Cancellation Penalties**:
  - Implement stricter penalties for cancellations made closer to the check-in date.
  - Offer flexible cancellation policies for a premium fee, allowing customers to opt for refundable bookings at an additional cost.
- **Encourage Non-Refundable Bookings**:
  - Provide discounts for non-refundable bookings to incentivize commitment.
  - Highlight savings for non-refundable options during the booking process.

Expected Impact:

- Discourages unnecessary cancellations.
- Generates revenue even for canceled bookings through penalty fees.

2. Implement Loyalty Programs

Loyal customers tend to cancel less frequently, making them a valuable segment to cultivate.

Key Insights Supporting This Recommendation:

- Repeated guests exhibit significantly lower cancellation rates compared to first-time customers.

Strategies:

- **Offer Exclusive Benefits**:
  - Reward repeat customers with perks like discounted rates, free upgrades, or complimentary meals.
  - Create a tiered loyalty program where benefits improve with frequency of stays.
- **Early Access to Bookings**:
  - Allow loyalty program members early access to popular room types or promotional deals.

Expected Impact:

- Enhances customer retention and encourages frequent stays.
- Builds a loyal customer base that generates consistent revenue.

3. Dynamic Pricing and Lead Time Management

Dynamic pricing strategies can optimize room revenue and reduce cancellations.

Key Insights Supporting This Recommendation:

- Bookings with longer lead times have higher cancellation rates.
- Room prices vary significantly across customer segments.

Strategies:

- **Adjust Prices Based on Lead Time**:
  - Offer discounts for bookings made closer to the check-in date to fill inventory.
  - Set slightly higher prices for long-lead-time bookings, offsetting the risk of cancellations.
- **Market-Segment-Specific Pricing**:
  - Offer tailored pricing for corporate and direct bookings, which tend to have lower cancellation rates.
  - Analyze competitive pricing in online travel agent segments and adjust dynamically to attract committed customers.

Expected Impact:

- Maximizes revenue during high-demand periods.
- Reduces inventory risk associated with long-lead-time bookings.

4. Optimize Resource Allocation

Predicting cancellations enables better resource allocation and planning.

Key Insights Supporting This Recommendation:

- High cancellation rates lead to unutilized rooms, impacting staffing and resource planning.

Strategies:

- **Overbooking Strategy**:
    - Use machine learning predictions to strategically overbook rooms in segments with historically high cancellation rates.
- **Staff Scheduling Optimization**:
    - Align staff schedules with predicted room occupancy to minimize idle resources.
- **Real-Time Monitoring**:
    - Use a dashboard to monitor booking trends and cancellations, allowing dynamic adjustments to operations.

Expected Impact:

- Improves operational efficiency.
- Reduces costs associated with underutilized resources.

5. Enhance Customer Engagement

Customers with special requests are less likely to cancel, highlighting the importance of engagement.

Key Insights Supporting This Recommendation:

- Guests with higher numbers of special requests exhibit lower cancellation rates.

Strategies:

- **Proactive Communication**:
    - Send personalized follow-ups or reminders to guests before their stay to confirm their booking and address any concerns.
- **Streamline Special Requests Handling**:
    - Ensure a seamless process for fulfilling special requests, increasing customer satisfaction and commitment.
- **Targeted Campaigns for High-Risk Segments**:

    - Identify and engage with customers from high-cancellation segments through offers or personalized messages.

Expected Impact:

- Enhances customer satisfaction and reduces cancellations.
- Builds trust and strengthens the hotel's brand equity.

6. Leverage Predictive Insights

The machine learning model provides actionable insights for cancellation prediction and risk management.

Key Insights Supporting This Recommendation:

- Predictive models achieve significant accuracy in identifying potential cancellations.

Strategies:

- **Cancellation Risk Flags**:
    - Implement a flagging system to alert staff of high-risk bookings based on model predictions.
- **Targeted Retention Efforts**:
    - Focus retention efforts (e.g., personalized offers or flexible booking options) on flagged bookings.
- **Automated Dynamic Adjustments**:
    - Automatically adjust availability and pricing based on predicted cancellations to ensure optimal inventory utilization.

Expected Impact:

- Enables proactive management of cancellations.
- Improves room occupancy and profitability.

7. Marketing and Promotions

Marketing efforts should focus on reducing cancellations and attracting committed bookings.

Key Insights Supporting This Recommendation:

- Certain segments, such as online travel agents, show higher cancellation rates.

Strategies:

- **Promote Direct Bookings**:
    - Encourage customers to book directly through the hotel's website by offering discounts or additional perks.

- **Segment-Specific Campaigns**:
    - Tailor promotions to segments with lower cancellation rates, such as corporate or repeated guests.
- **Seasonal Campaigns**:
    - Offer special packages during low-demand periods to attract bookings.

Expected Impact:

- Reduces dependency on high-cancellation segments.
- Attracts a diverse and reliable customer base.

Conclusion

By combining data-driven insights, predictive analytics, and targeted business strategies, INN Hotels Group can significantly reduce cancellations and improve profitability. Implementing these actionable recommendations will enhance operational efficiency, foster customer loyalty, and position the hotel group for sustainable growth in a competitive hospitality market.

Conclusion

The challenges posed by booking cancellations are multifaceted but not insurmountable. By leveraging machine learning models, INN Hotels can predict cancellations with significant accuracy and take proactive measures to mitigate revenue loss.

The findings of this report provide a roadmap for implementing predictive analytics, revising business policies, and enhancing customer engagement. INN Hotels Group stands to benefit immensely by transforming the insights into strategic actions, fostering customer loyalty, and improving operational efficiency. With this data-driven approach, the organization is poised to achieve sustainable growth while navigating the dynamic demands of the hospitality industry.