Unlocking Hospitality Insights: A Data-Driven Exploration

ABSTRACT

This study aims to utilize a rich hotel booking dataset to uncover valuable patterns and trends in the hospitality industry through machine learning and data analysis. The goal is to extract insights that can optimize hotel management strategies, from understanding guest preferences to maximizing revenue. This data-driven approach has the potential to significantly improve guest experiences and propel the hotel sector forward.

PRESENTED BY:

Group No 7

Vishal Tiwari 2304868

Naman Khandelwal 2309377

Hrishikesh Kadam 2302167

PRESENTED TO:

Dr. Shashikant Deepak

Unlocking Hospitality Insights: A Data-Driven Exploration

Executive Summary:

We are driven to venture into the field of hotel management by the prospect of data-driven discoveries. We are concentrating on the 'hotel booking' dataset from Kaggle, which is a veritable gold mine of data with 119,390 entries and 31 characteristics related to City and Resort hotels. Our mission is to use machine learning and data analysis to find useful patterns, trends, and tactics in this dataset. Our goal is to extract useful information that can improve hotel management techniques and have a significant impact on the hospitality sector, from identifying client preferences to maximizing income. Come along on this adventure as we uncover the possibilities hidden in hotel booking data, with the hope that data-driven choices can improve visitor experiences and propel the hotel sector forward.

Problem Statement:

How can a hotel use data analytics to increase revenue, lower cancellation rates, improve booking management, and improve guest satisfaction?

Utilizing the 'hotel booking' dataset to build a thorough analytical model that would forecast the possibility of cancellations and optimize room price schemes. Predictive analytics will be incorporated into the model to forecast cancellations based on seasonal trends, booking patterns, and client profiles. To further identify unique client groups and their preferences, the investigation will also incorporate customer segmentation. Utilizing these insights to develop dynamic pricing strategies that optimize revenue and sustain high occupancy rates is the ultimate goal. This concept will give the hotel a competitive edge in the industry and help to improve its revenue management system.

The primary objective of this analysis is to harness the power of data analytics and machine learning to optimize hotel booking management, reduce cancellation rates, maximize revenue, and enhance guest satisfaction. To achieve this, we will explore various facets of the dataset and address the following key questions:

1. **Customer Demographics and Preferences**: We will investigate the geographical distribution of guests, customer types, room preferences, and special requests to tailor services effectively.

2. **Seasonal Trend Analysis:** By examining booking dates and lead times, we will identify peak and off-peak seasons, aiding in pricing and promotion strategies.

3. **Customer Segmentation Cancellation Analytics:** Understanding cancellation rates and identifying factors correlated with higher cancellations will inform policy adjustments and revenue forecasting.

4. **Revenue Management**: We will analyze average daily rates (ADR) across different times of the year and customer segments, helping to formulate dynamic pricing strategies and resource allocation.

5. **Predictive Analytics:** Utilizing historical data, we will build predictive models to forecast future bookings and cancellations, enabling better capacity planning and staffing.

Data Collection and Preparation:

The first step in our research was to use Pandas to load the 'hotel_booking.csv' dataset. We looked at column names and types briefly before addressing missing values. The 'agent' and 'business' columns were eliminated due to large gaps in them. The mode was used to substitute missing values for "children," and missing country data was standardized to "Unknown." These precautions guaranteed the integrity of the dataset, which made it possible for us to analyze hotel reservation insights later on.

Data set Link: https://www.kaggle.com/datasets/jessemostipak/hotel-booking-demand/data

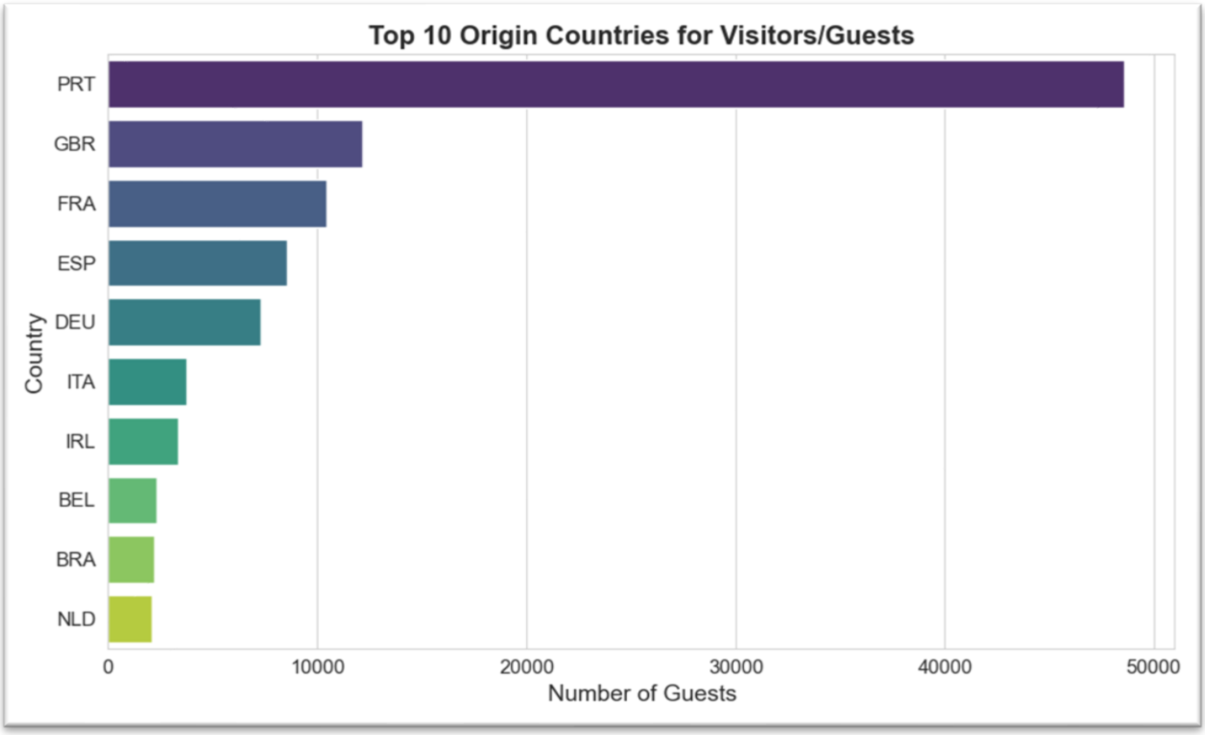
Data Analysis:

1. Customer Demographics and Preferences

Our analysis reveals compelling insights into the demographics and preferences of hotel guests. This section delves into the geographical origins of guests, their types, room preferences, and special requests – each of which is critical for tailoring services and enhancing customer satisfaction.

Geographical Distribution of Guests

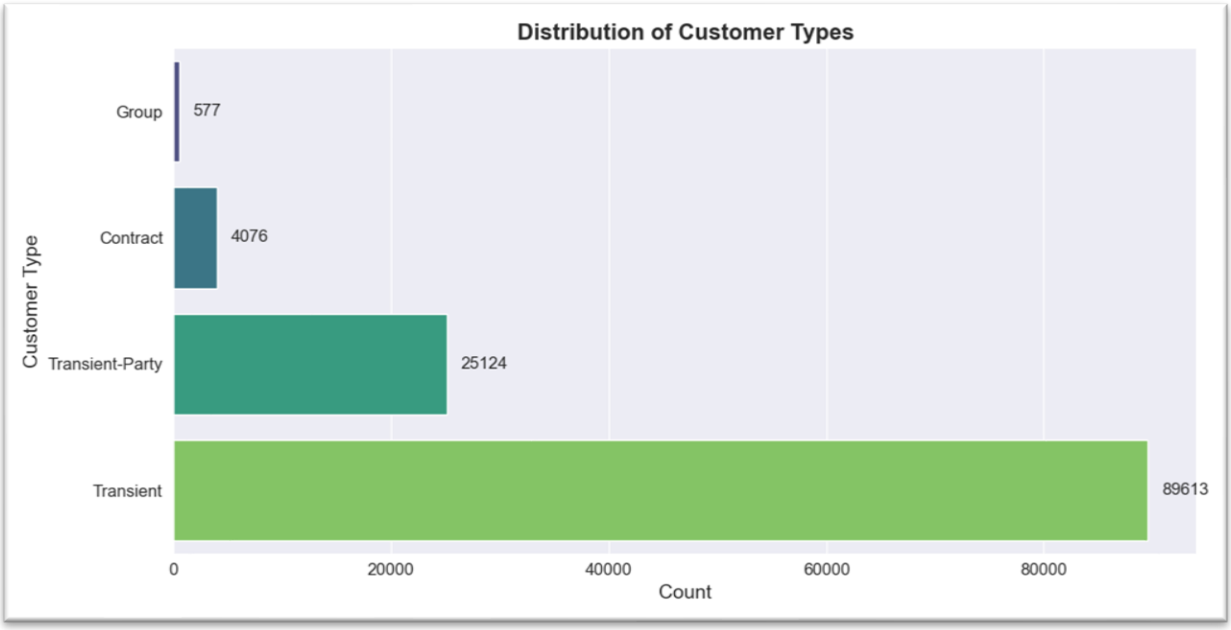
The first visualization provides a clear indication of the international diversity of the hotel's clientele. The bar chart titled "Top 10 Countries of Origin for Guests" shows that the majority of the guests originate from Portugal, followed by Great Britain, and France. This data suggests that targeted marketing strategies in these regions could be highly effective. For country codes refer: https://countrycode.org/

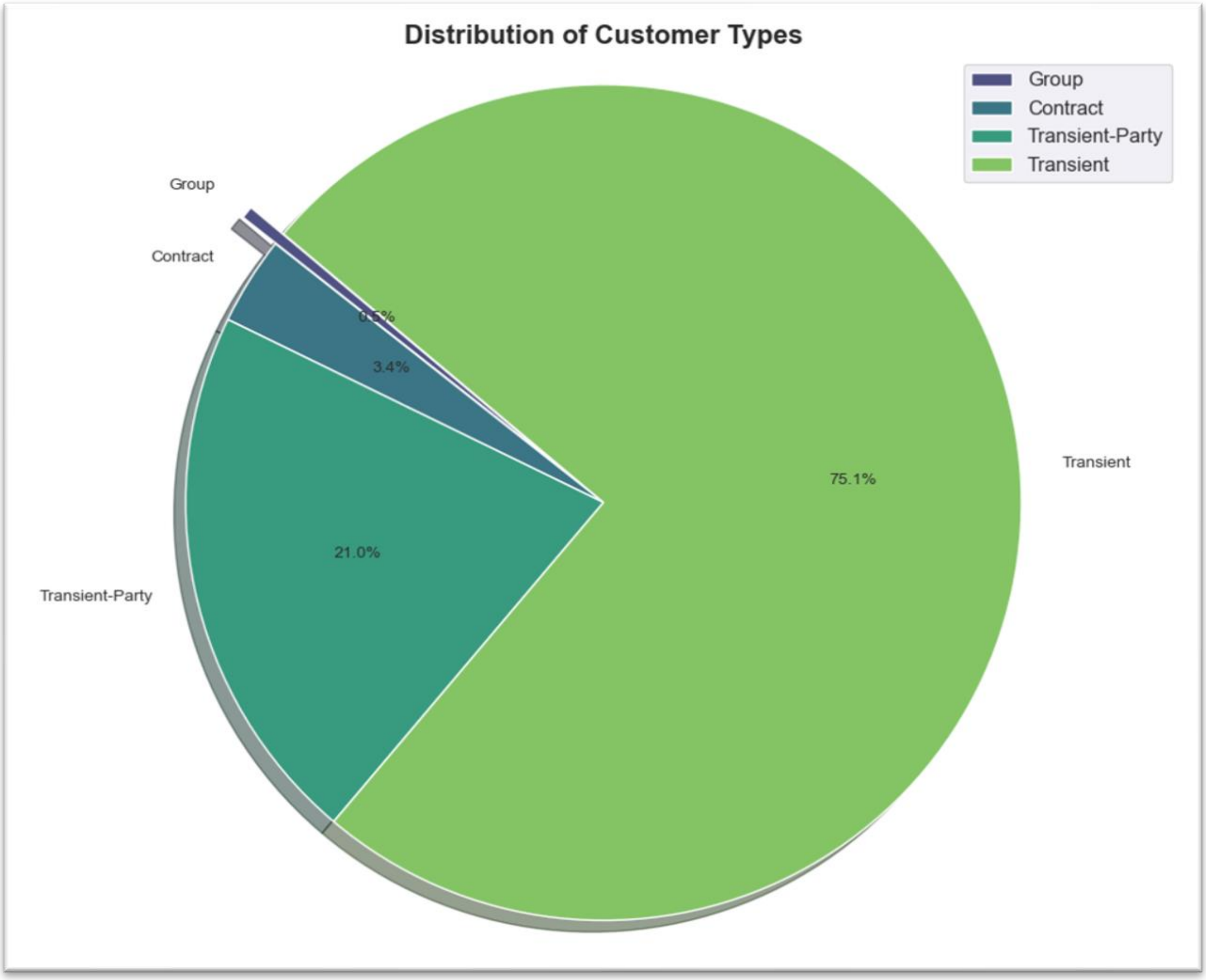


Geographical Distribution of Guests

Customer Types

Turning now to customer types, the bar and pie charts show that 'Transient' clients make up the largest sector, making up an astounding 75.1% of all customers. The next group of consumers is 'Transient-Party,' accounting for 21%; 'Contract' and 'Group' customers follow with 3.4% and 0.5%, respectively. This suggests that our customers travel primarily alone or in small groups, which may lead us to concentrate on customization and unique visitor experiences.

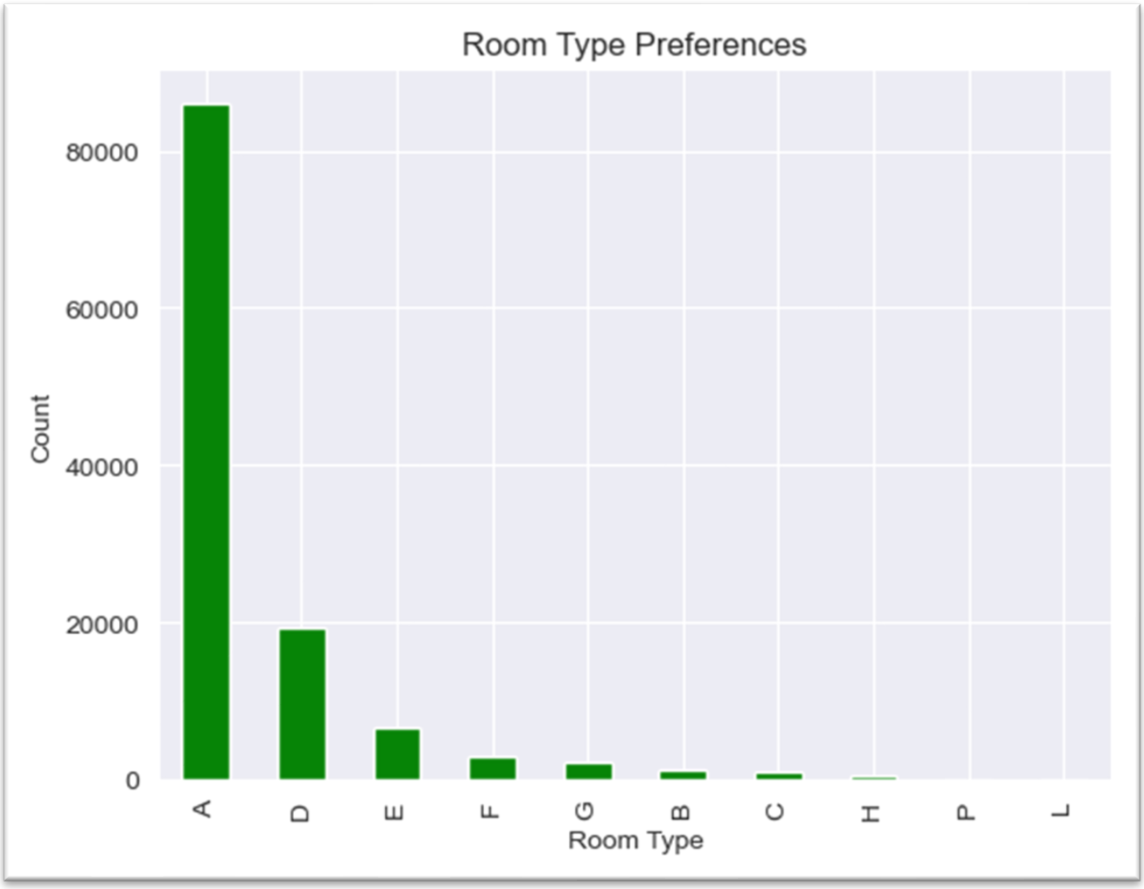




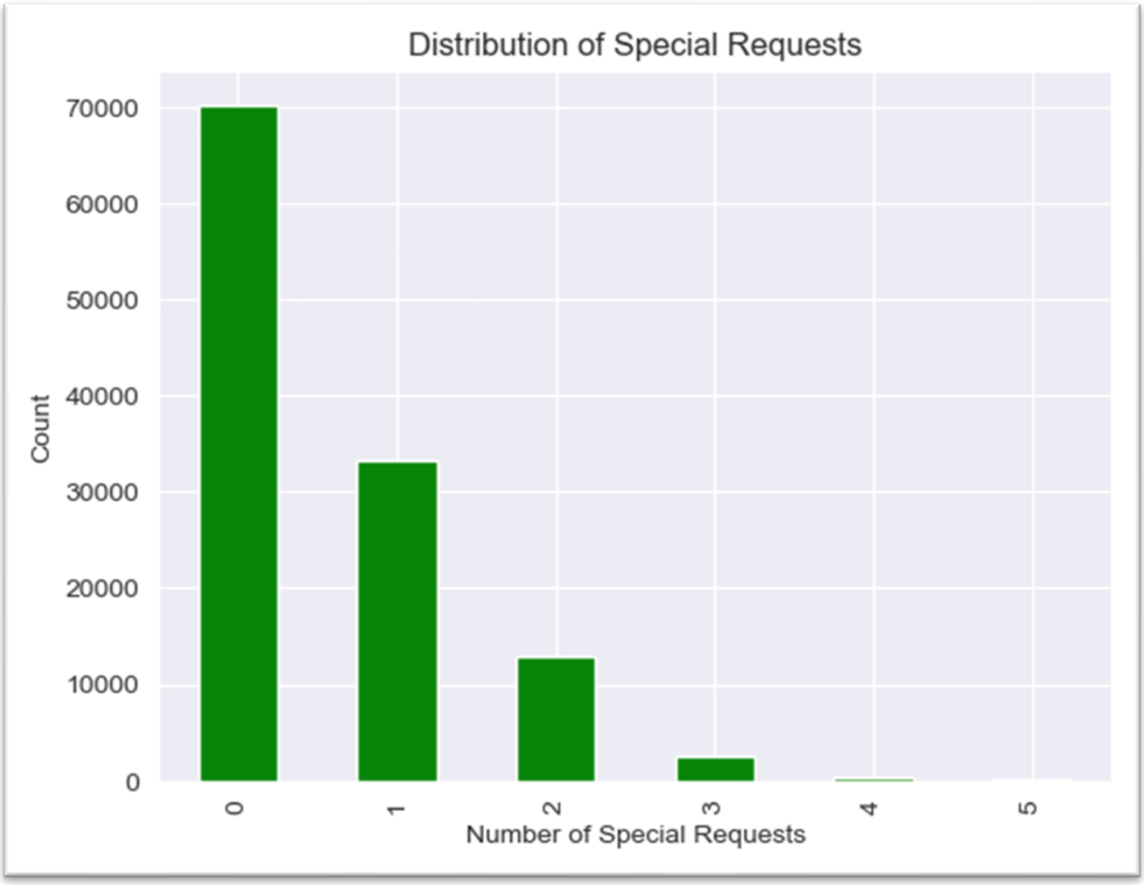
Distribution of Customer_types

Room Preferences and Special Requests

The information on special requests and preferred rooms paints an accurate picture of the priorities of the visitors. A clear tendency toward a certain room type, which emerges as the most frequently booked, can be seen in our examination of room type preferences. A major contributor to increasing visitor happiness and promoting return business may be guaranteeing the availability and upholding the high standards of these chosen rooms. In addition, a sizable portion of visitors choose not to seek extra services, according to the special requests statistics. This data points to an unexplored possibility for actively promoting these offerings, which could improve the visitor experience and encourage loyalty. We may incrementally boost revenue while also personalizing our service by emphasizing the option of customizing their stay through specific requests.



Room_Type_Preference



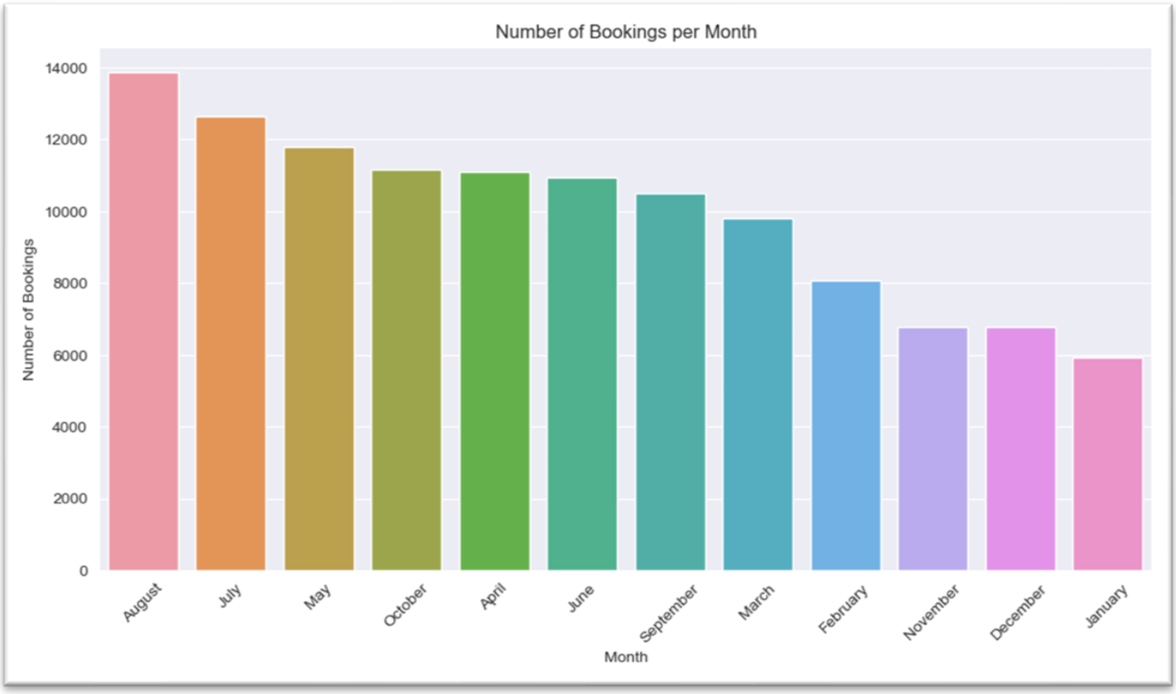
Distribution of Special_Requests

2. Analysis of Seasonal Trends:

Our analysis of lead times and booking dates has revealed some clear trends that are essential for creating efficient pricing and marketing plans.

Seasons of Peak and Off-Peak :

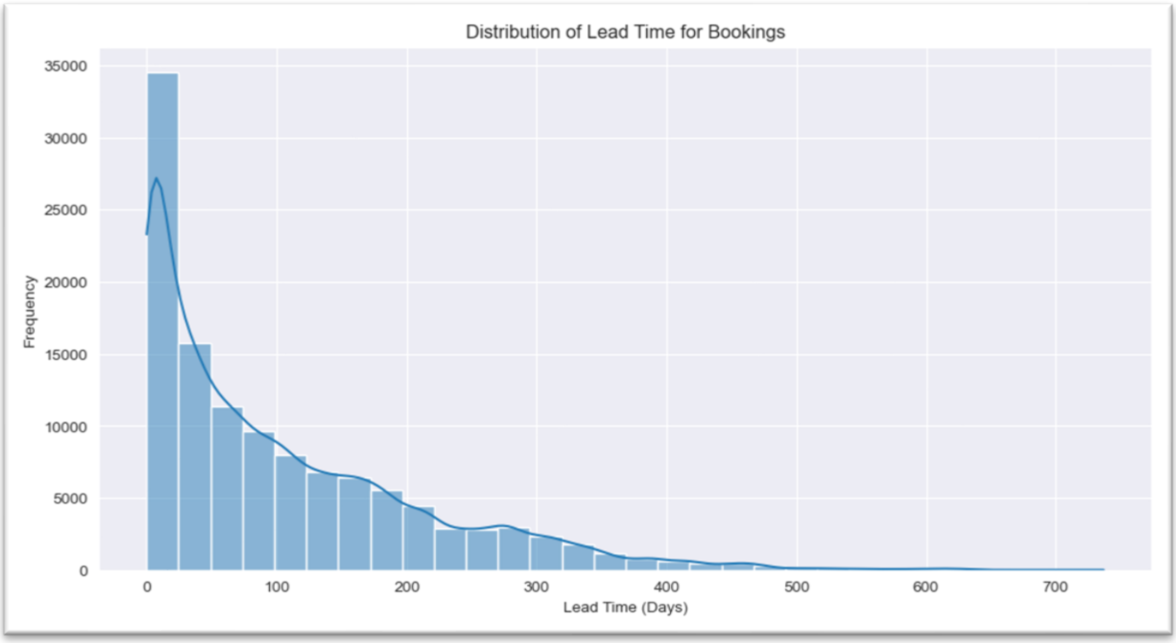
The "Number of Bookings per Month" bar graph shows the annual fluctuations in the number of guest reservations. August, July, and May are clearly the months with the most bookings, with August having the most. Due to the summer holiday season, when travelers are more likely to travel, this predicts a strong demand for lodging during these summer months. On the other hand, the months with lower booking numbers—January, December, and November—represent off-peak times when the hotel may run marketing initiatives to raise occupancy rates.



Number_of_Bookings_per_Month

Lead Times for Bookings

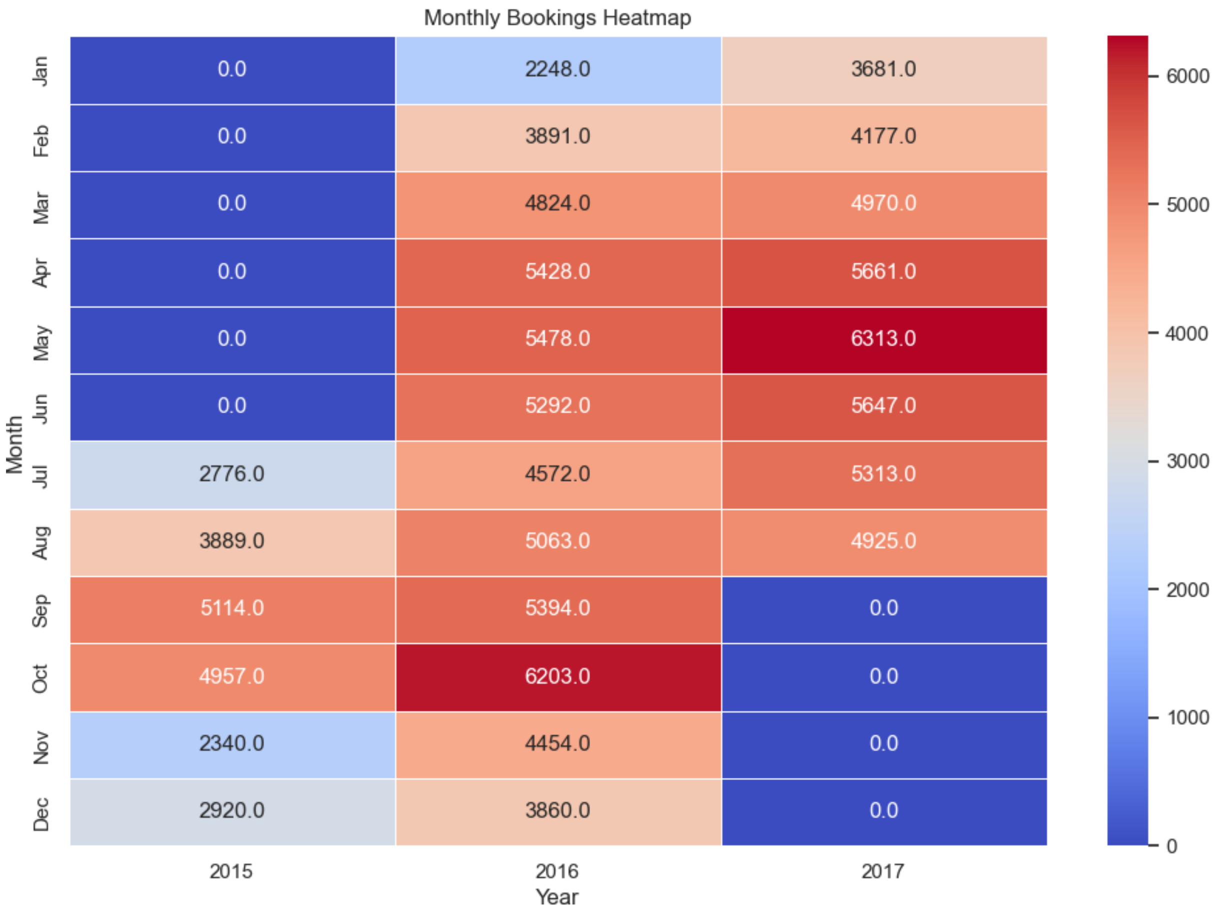
A considerable proportion of appointments are made with a comparatively short advance time, peaking at 0–10 days, according to the "Distribution of Lead Time for Bookings" histogram. The significance of last-minute booking tactics and the opportunity to seize this market with timely offers are underscored by this research. The histogram does, however, also reveal a tail of bookings that extends towards longer lead periods, highlighting the need for early-bird discounts that target travelers who make reservations far in advance.



Distribution_of_Lead_Time_for_Bookings

Annual Reservation Trends

The "Monthly Bookings Heatmap" is a heatmap that shows both regular trends and anomalies when compared year over year. For example, for every year displayed, there is a discernible increase in bookings from June to October. The largest concentrations were detected in October 2016 and May 2016, indicating periods of peak demand. The heatmap also reveals a striking lack of data for the months of September to December in 2017 and January to June in 2015. This might be because the hotel was closed or didn't take reservations during those times, or it could be because the dataset didn't cover these months. We can not only forecast demand with this comparison view, but we can also spot anomalous variations that may help us with our marketing and operational readiness.



Monthly_Booking_Heat_Map

3. Analytics for Customer Segmentation and Cancellation:

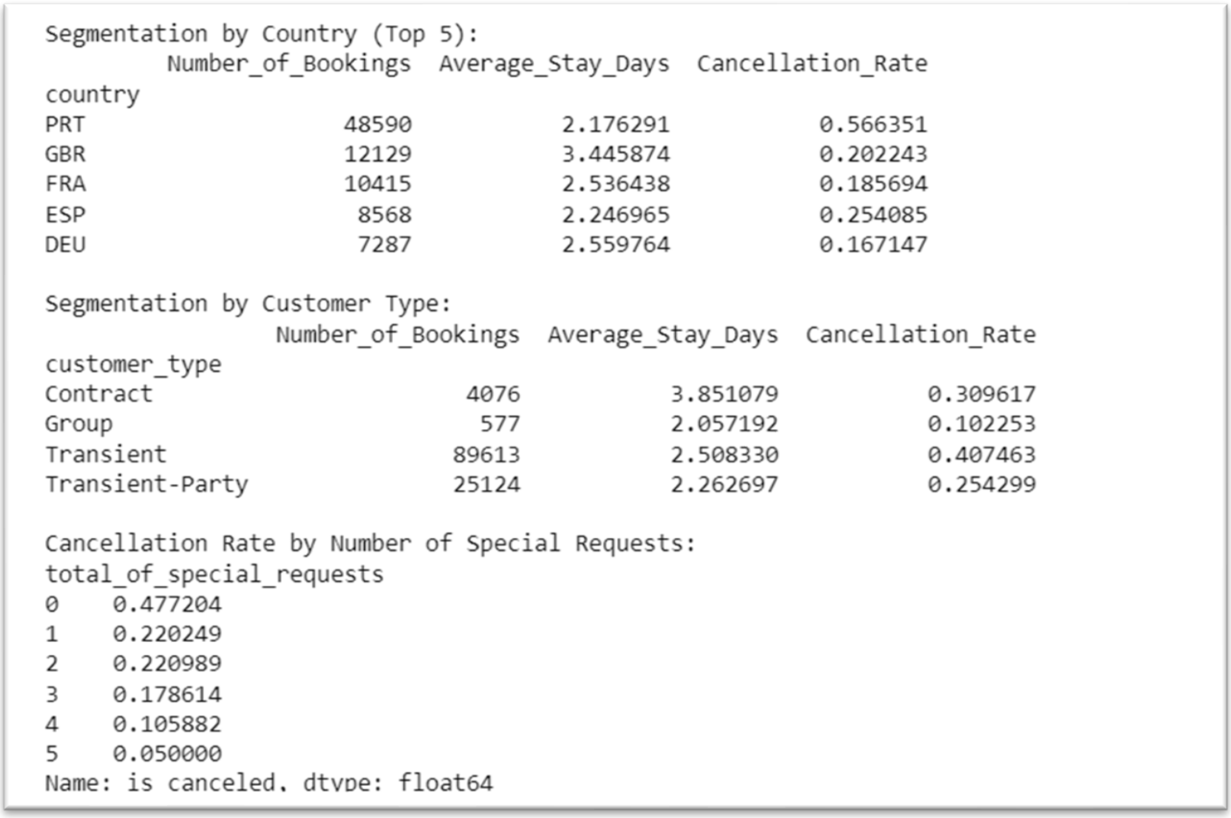
a. Transient clients, or those who are just visiting for a short time, have the greatest cancellation rate—roughly 41%.

b. A booking's cancellation rate decreases with the number of specific requests made. The cancellation rate for reservations without any special requests is roughly 47.72%, but the rate for reservations with five special requests is only 5.00%.

c. There is a strong correlation between cancellation rates and the quantity of bookings. Higher booking volume typically translates into higher cancellation rates, and vice versa.

An Introduction to Customer Segmentation

When broken down by nation, Portuguese visitors have the most reservations out of the top five, but they also have the greatest cancellation rate. This can point to a cultural or market-specific trend where bookings are made frequently by local guests but are also more susceptible to change. On the other hand, guests from Germany show the lowest cancellation rate, which might reflect a more decisive booking behavior or better planning.



```
Segmentation by Country (Top 5):
         Number_of_Bookings  Average_Stay_Days  Cancellation_Rate
country
PRT                   48590           2.176291           0.566351
GBR                   12129           3.445874           0.202243
FRA                   10415           2.536438           0.185694
ESP                    8568           2.246965           0.254085
DEU                    7287           2.559764           0.167147

Segmentation by Customer Type:
                 Number_of_Bookings  Average_Stay_Days  Cancellation_Rate
customer_type
Contract                       4076           3.851079           0.309617
Group                           577           2.057192           0.102253
Transient                     89613           2.508330           0.407463
Transient-Party               25124           2.262697           0.254299

Cancellation Rate by Number of Special Requests:
total_of_special_requests
0    0.477204
1    0.220249
2    0.220989
3    0.178614
4    0.105882
5    0.050000
Name: is_canceled, dtype: float64
```

Segmentation_by_country_&_Customer_Type

When we categorize customers by type, 'Contract' customers have a lower cancellation rate than 'Transient' customers but higher than 'Group' customers. This might be due to the structured nature of contracts that come with penalties or non-refundable clauses for cancellations. 'Transient' customers, as previously noted, have a higher cancellation rate which aligns with their temporary travel nature. 'Group' bookings are the least likely to be canceled, possibly due to the logistical complexity and planning involved in group travel arrangements.

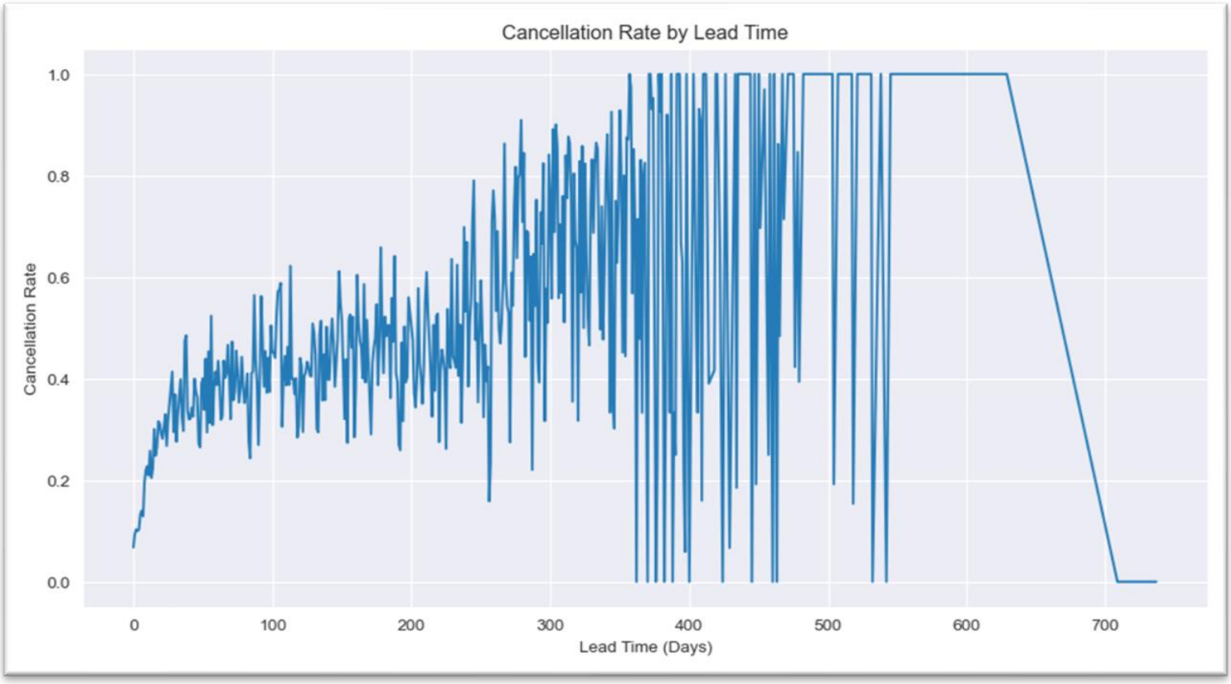
The data on special requests further confirms that personal investment in a booking correlates with a lower cancellation rate. It is noteworthy that each additional special request seems to significantly reduce the likelihood of cancellation, indicating that personalized service is not only a key to guest satisfaction but also to securing the booking.

Lead Time Analysis

The analysis reveals a connection between longer booking lead times (time between booking and arrival) and higher cancellation rates. This means that customers booking further in advance tend to cancel more often.

This information can be valuable for:

- **Understanding customer segments:** We can analyze which types of customers book further in advance and are more likely to cancel.
- **Reducing cancellations:** By understanding the factors influencing cancellations, we can develop targeted strategies to mitigate them.
- **Optimizing booking policies:** We can adjust booking policies based on the lead time-cancellation rate relationship.



Cancellation_rate_by_lead_time

Correlation Heat Map

The heatmap is a visualization tool used to understand the relationships between multiple variables at once, in this case, within a hotel booking dataset. Here's an overall interpretation of the heatmap:

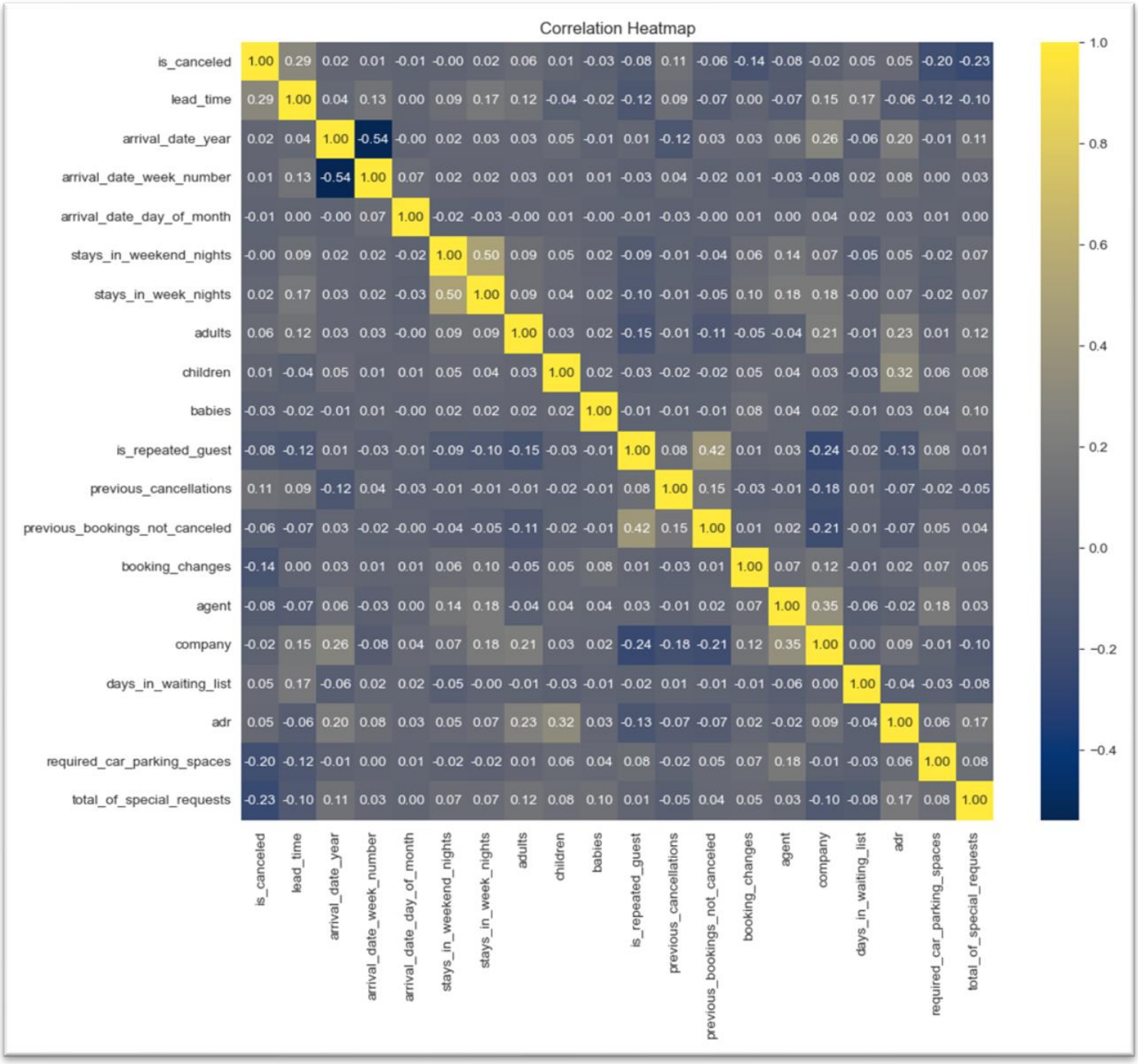
Strong Positive Correlations: Certain variables, like lead_time and previous_cancellations, show a strong positive correlation with the is_canceled variable. This indicates that bookings made far in advance or by customers who have previously canceled are more likely to be canceled again. Such patterns suggest that early bookings and customers' cancellation histories could be reliable indicators of potential future cancellations.

Strong Negative Correlations: The variables total_of_special_requests and required_car_parking_spaces have strong negative correlations with is_canceled. This suggests that when customers make more special requests or need parking spaces, they are less likely to cancel their booking. This could be due to a higher level of commitment to the stay when specific needs are expressed or planned for.

Weak or No Correlations: Several variables such as arrival_date_year, arrival_date_month, and babies show weak correlations with the is_canceled variable, indicating that these factors have little to no linear relationship with the likelihood of a booking being canceled. This means that simply knowing the year or month of booking, or whether babies are included in the party, doesn't provide strong predictive power about cancellations.

Correlation among Independent Variables: Aside from the target variable is_canceled, the heatmap also shows how independent variables relate to each other. For example, adr (Average Daily Rate) shows a positive correlation with the number of adults and children, which implies that bookings for larger groups tend to be associated with a higher rate.

Overall, the heatmap is a strategic tool that can guide hotel management in identifying risk factors for cancellations, understanding customer behavior, and making informed decisions about hotel policies and marketing strategies. By analyzing these correlations, management can target interventions to mitigate cancellation risks and tailor services to customer needs, ultimately aiming to improve occupancy rates and revenue.



	is_canceled	lead_time	arrival_date_year	arrival_date_week_number	arrival_date_day_of_month	stays_in_weekend_nights	stays_in_week_nights	adults	children	babies	is_repeated_guest	previous_cancellations	previous_bookings_not_canceled	booking_changes	agent	company	days_in_waiting_list	adr	required_car_parking_spaces	total_of_special_requests
is_canceled	1.00	0.29	0.02	0.01	-0.01	-0.00	0.02	0.06	0.01	-0.03	-0.08	0.11	-0.06	-0.14	-0.08	-0.02	0.05	0.05	-0.20	-0.23
lead_time	0.29	1.00	0.04	0.13	0.00	0.09	0.17	0.12	-0.04	-0.02	-0.12	0.09	-0.07	0.00	-0.07	0.15	0.17	-0.06	-0.12	-0.10
arrival_date_year	0.02	0.04	1.00	-0.54	-0.00	0.02	0.03	0.03	0.05	-0.01	0.01	-0.12	0.03	0.03	0.06	0.26	-0.06	0.20	-0.01	0.11
arrival_date_week_number	0.01	0.13	-0.54	1.00	0.07	0.02	0.02	0.03	0.01	0.01	-0.03	0.04	-0.02	0.01	-0.03	-0.08	0.02	0.08	0.00	0.03
arrival_date_day_of_month	-0.01	0.00	-0.00	0.07	1.00	-0.02	-0.03	-0.00	0.01	-0.00	-0.01	-0.03	-0.00	0.01	0.00	0.04	0.02	0.03	0.01	0.00
stays_in_weekend_nights	-0.00	0.09	0.02	0.02	-0.02	1.00	0.50	0.09	0.05	0.02	-0.09	-0.01	-0.04	0.06	0.14	0.07	-0.05	0.05	-0.02	0.07
stays_in_week_nights	0.02	0.17	0.03	0.02	-0.03	0.50	1.00	0.09	0.04	0.02	-0.10	-0.01	-0.05	0.10	0.18	0.18	-0.00	0.07	-0.02	0.07
adults	0.06	0.12	0.03	0.03	-0.00	0.09	0.09	1.00	0.03	0.02	-0.15	-0.01	-0.11	-0.05	-0.04	0.21	-0.01	0.23	0.01	0.12
children	0.01	-0.04	0.05	0.01	0.01	0.05	0.04	0.03	1.00	0.02	-0.03	-0.02	-0.02	0.05	0.04	0.03	-0.03	0.32	0.06	0.08
babies	-0.03	-0.02	-0.01	0.01	-0.00	0.02	0.02	0.02	0.02	1.00	-0.01	-0.01	-0.01	0.08	0.04	0.02	-0.01	0.03	0.04	0.10
is_repeated_guest	-0.08	-0.12	0.01	-0.03	-0.01	-0.09	-0.10	-0.15	-0.03	-0.01	1.00	0.08	0.42	0.01	0.03	-0.24	-0.02	-0.13	0.08	0.01
previous_cancellations	0.11	0.09	-0.12	0.04	-0.03	-0.01	-0.01	-0.01	-0.02	-0.01	0.08	1.00	0.15	-0.03	-0.01	-0.18	0.01	-0.07	-0.02	-0.05
previous_bookings_not_canceled	-0.06	-0.07	0.03	-0.02	-0.00	-0.04	-0.05	-0.11	-0.02	-0.01	0.42	0.15	1.00	0.01	0.02	-0.21	-0.01	-0.07	0.05	0.04
booking_changes	-0.14	0.00	0.03	0.01	0.01	0.06	0.10	-0.05	0.05	0.08	0.01	-0.03	0.01	1.00	0.07	0.12	-0.01	0.02	0.07	0.05
agent	-0.08	-0.07	0.06	-0.03	0.00	0.14	0.18	-0.04	0.04	0.04	0.03	-0.01	0.02	0.07	1.00	0.35	-0.06	-0.02	0.18	0.03
company	-0.02	0.15	0.26	-0.08	0.04	0.07	0.18	0.21	0.03	0.02	-0.24	-0.18	-0.21	0.12	0.35	1.00	0.00	0.09	-0.01	-0.10
days_in_waiting_list	0.05	0.17	-0.06	0.02	0.02	-0.05	-0.00	-0.01	-0.03	-0.01	-0.02	0.01	-0.01	-0.01	-0.06	0.00	1.00	-0.04	-0.03	-0.08
adr	0.05	-0.06	0.20	0.08	0.03	0.05	0.07	0.23	0.32	0.03	-0.13	-0.07	-0.07	0.02	-0.02	0.09	-0.04	1.00	0.06	0.17
required_car_parking_spaces	-0.20	-0.12	-0.01	0.00	0.01	-0.02	-0.02	0.01	0.06	0.04	0.08	-0.02	0.05	0.07	0.18	-0.01	-0.03	0.06	1.00	0.08
total_of_special_requests	-0.23	-0.10	0.11	0.03	0.00	0.07	0.07	0.12	0.08	0.10	0.01	-0.05	0.04	0.05	0.03	-0.10	-0.08	0.17	0.08	1.00

Correlation_Heatmap

#Plasma: A colormap with vibrant colors that smoothly transitions through a variety of shades.

#Inferno: Similar to Plasma but with a different color distribution, often used for visualizations of data with a darker background.

#Magma: A colormap that emphasizes high-intensity colors, often used for emphasizing areas of high value in data.

#Cividis: Designed to be perceptually uniform and suitable for users with color vision deficiencies.

#Turbo: A colormap with vibrant and distinct colors, often used for highlighting different data values.

Plasma and Inferno:

- These are both visually appealing colormaps with smooth transitions and vibrant colors.
- Plasma offers a wider range of hues, while Inferno leans towards a more red-orange-yellow spectrum.
- Both are suitable for data visualizations where you want to emphasize the magnitude of values, often used for continuous data ranging from low to high.

Magma:

- This colormap prioritizes high-intensity colors, making it ideal for highlighting areas of high data values.
- It's particularly useful when you want to draw attention to specific regions within your visualization.

Cividis:

- This colormap is specifically designed to be colorblind-friendly and perceptually uniform.
- It uses a carefully chosen sequence of colors that are distinguishable even for individuals with color vision deficiencies.
- Cividis is a great choice when you want to ensure your visualizations are accessible to a wider audience.

Turbo:

- This colormap features distinct and vibrant colors, making it suitable for highlighting different data values with clear separation.
- It's often used for categorical data where you want to visually distinguish between different categories.

Choosing the right colormap depends on the specific type of data you are visualizing and the message you want to convey. Here's a quick summary:

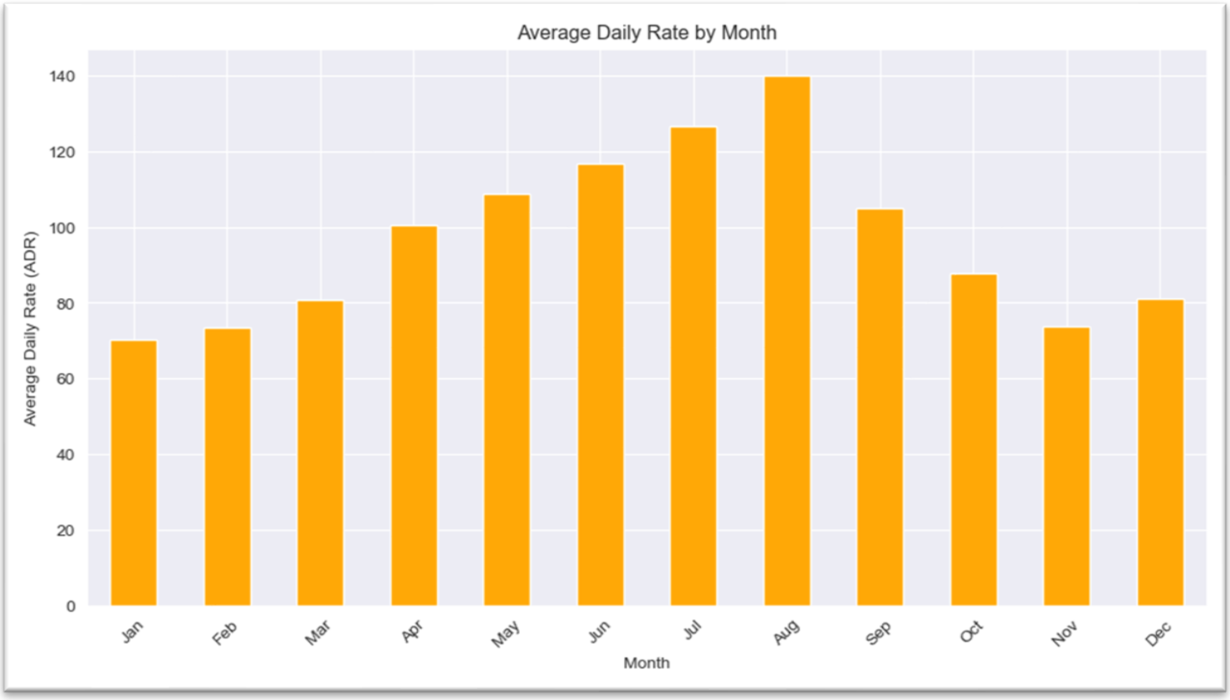
- **For continuous data with emphasis on magnitude:** Plasma, Inferno
- **For highlighting high-value areas:** Magma
- **For colorblind-friendly visualizations:** Cividis
- **For distinguishing between categories:** Turbo

Remember, the key is to select a colormap that effectively communicates your data story and is accessible to your target audience.

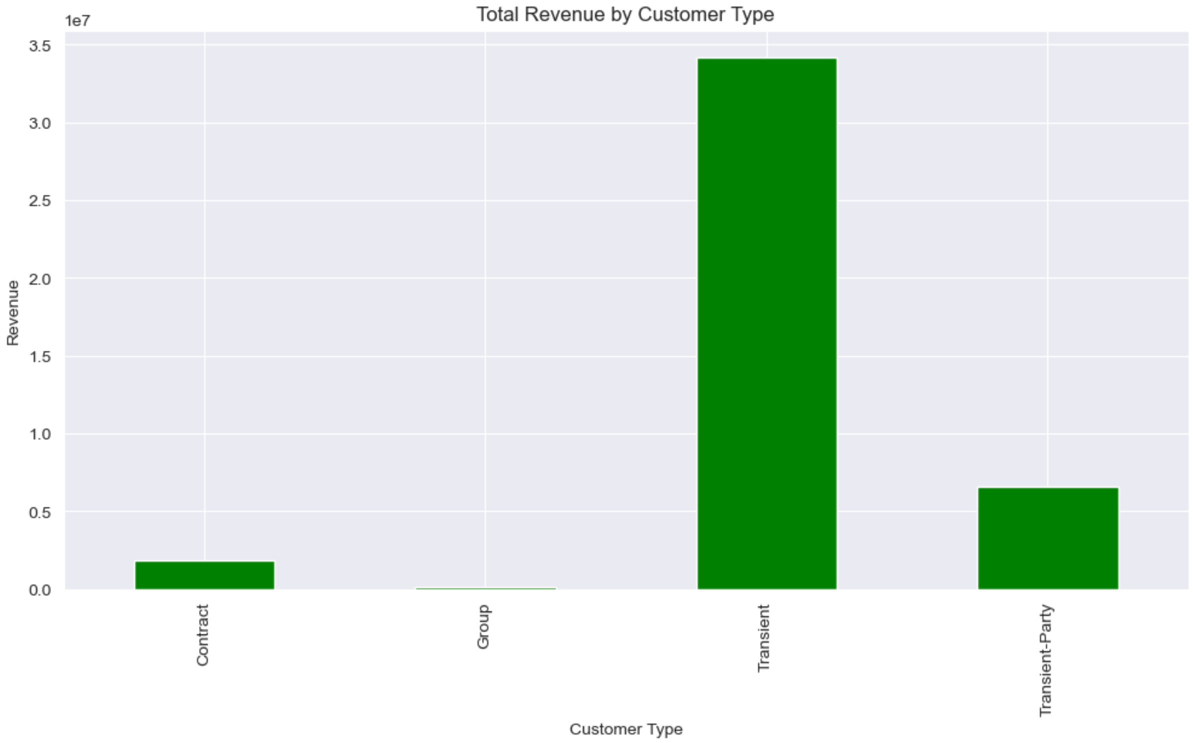
4. Revenue Management:

The bar chart depicting the Average Daily Rate (ADR) by month for hotel bookings reveals a seasonal pricing pattern, with peaks typically during the summer months of June, July, and August, suggesting higher demand possibly due to vacationers and favorable weather. In contrast, the ADR dips in the post-summer months, with the lowest rates observed at the beginning of the year, in January and February, likely reflecting a slowdown after the holiday season. This fluctuation in ADR underscores the importance of dynamic pricing strategies in the hotel industry to optimize revenue, where prices are elevated during high-demand periods and discounts or promotions might be applied during slower months to maintain occupancy rates.

Also, the second bar chart represents total revenue generated from different customer types at a hotel. It shows that the 'Transient' customer type contributes significantly more to revenue than 'Contract' or 'Transient-Party' types. The 'Group' category is not visible, suggesting minimal to no revenue contribution from this segment within the data's scope. This indicates that individual travelers, likely booking for short stays, are the primary revenue drivers for the hotel.



ADR by month



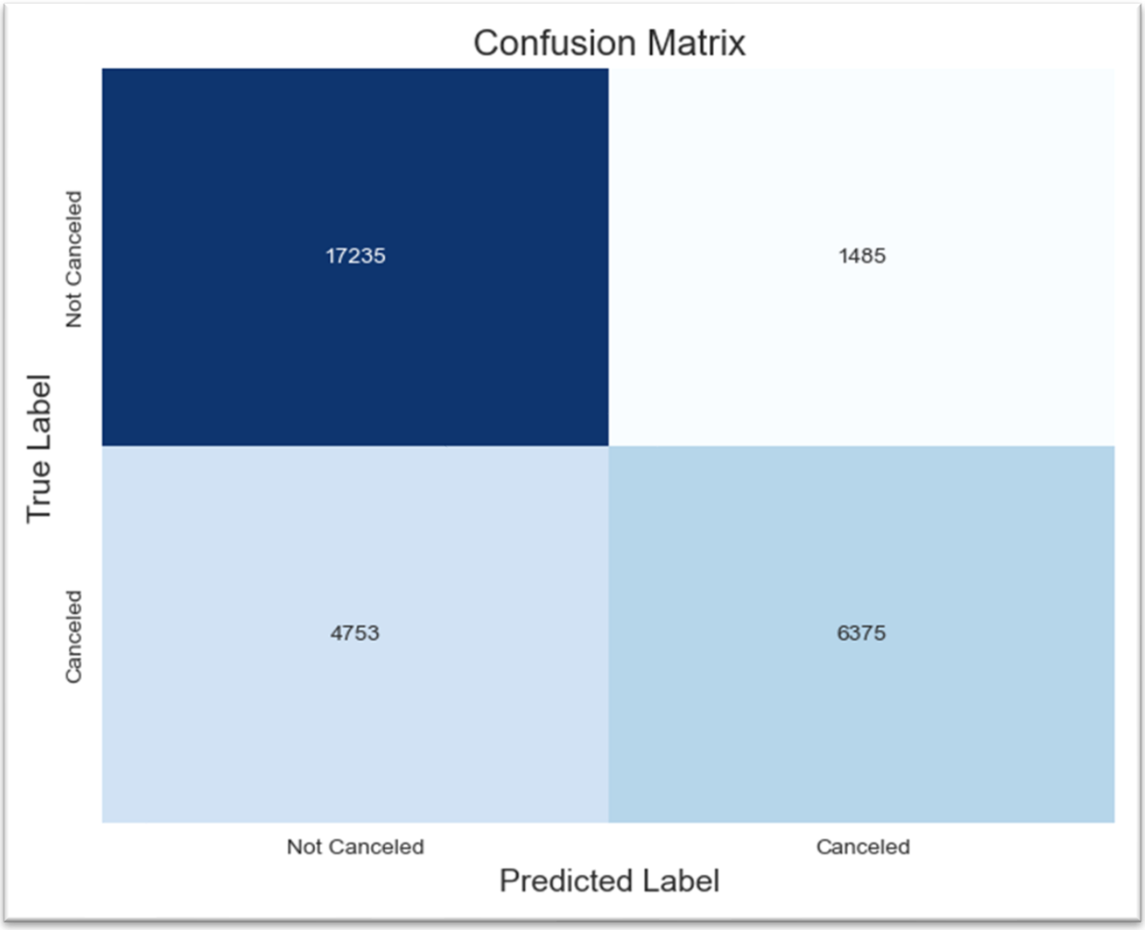
Total_revenue_by_customer_Type

5. Predictive Analytics:

Classification Report

In this Logistic regression model For class 0 (likely representing bookings that were not canceled):
Precision: 78% (When the model predicts a booking will not be canceled, it is correct 78% of the time.)
Recall: 92% (The model correctly identifies 92% of all actual non-canceled bookings.)
F1-score: 85% (A measure of the test's accuracy, combining precision and recall for class 0.)
Support: 18,720 (The number of actual non-canceled bookings in the test set.)
For class 1 (likely representing bookings that were canceled):
Precision: 81% (When the model predicts a booking will be canceled, it is correct 81% of the time.)
Recall: 57% (The model correctly identifies 57% of all actual canceled bookings.) F1-score: 67% (A measure of the test's accuracy for class 1.)
The model's overall accuracy is 79%, indicating that it correctly predicts 79% of the booking cancellations across both classes.

Confusion Matrix: The matrix is a 2x2 table that reports the number of true positives, false negatives, false positives, and true negatives.
For non-canceled bookings (class 0): 17,235 were correctly predicted (true positives), and 1,485 were wrongly predicted as canceled (false negatives).
For canceled bookings (class 1): 4,753 were wrongly predicted as non-canceled (false positives), and 6,375 were correctly predicted as canceled (true negatives).

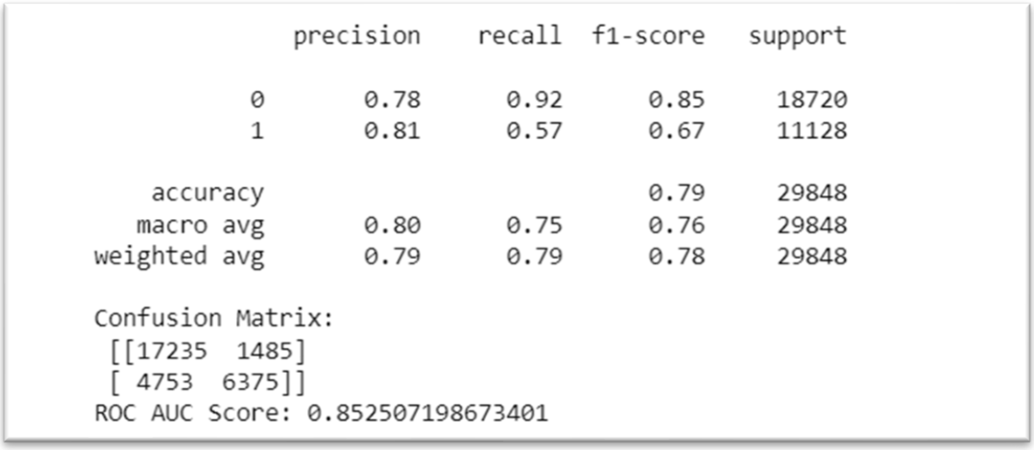


Confusion Matrix

```
ROC AUC Score: 0.852507198673401
```

The ROC AUC Score of 0.8525 indicates the model's ability to distinguish between the classes is very good. The score ranges from 0.5 (no better than random chance) to 1 (perfect prediction), and a score above 0.8 is considered to be very good.

In summary, the logistic regression model is quite good at predicting whether a booking will be canceled or not, with particularly strong performance in identifying the non-canceled bookings. However, it is less effective in correctly identifying the canceled bookings, as evidenced by the lower recall for class 1. The ROC AUC score suggests that the model has a good measure of separability and is capable of distinguishing between canceled and non-canceled bookings effectively.



```
              precision    recall  f1-score   support

           0       0.78      0.92      0.85     18720
           1       0.81      0.57      0.67     11128

    accuracy                           0.79     29848
   macro avg       0.80      0.75      0.76     29848
weighted avg       0.79      0.79      0.78     29848

Confusion Matrix:
 [[17235  1485]
 [ 4753  6375]]
ROC AUC Score: 0.852507198673401
```

Solution:

- ### Seasonal Trend Insights:

➔ Optimize Pricing: Increase prices in peak summer months, offer discounts or packages in off-peak times.

➔ Staffing and Resource Allocation: Adequate staff during busy season, reduced during slow months.

➔ Maintenance Scheduling: Conduct renovations in off-peak periods to minimize impact on guests and revenue.

➔ Marketing Campaigns: Create campaigns for low seasons, like special winter packages or local event promotions.

- ### Short Lead Time Bookings:

➔ Last-Minute Deals: Offer incentives for last-minute bookings.

➔ Flexible Policies: Implement flexible cancellation/rebooking policies.

➔ Targeted Marketing: Use ads and promotions for spontaneous travelers.

- ### Advance Bookings:

➔ Early-Bird Specials: Discounts or perks for early bookings.

➔ Loyalty Programs: Benefits for repeat customers booking early.

- Lead Time Distribution (0-10 Days):

➔ Maintain Flexibility: Reserve rooms for last-minute bookings.
➔ Dynamic Pricing: Adjust rates based on inventory and booking date.
➔ Promotional Activities: Promote last-minute deals, especially online.

- Long-Tail Distribution:

➔ Early Booking Incentives: Offer advantages for far-in-advance reservations.
➔ Predictive Planning: Use data for revenue management and operational adjustments.

- Heat Map Analysis:

➔ Understanding Seasonality: Plan marketing, staffing, and maintenance based on busy/quiet months.
➔ Spotting Trends and Anomalies: Identify changes in booking patterns year-over-year.
➔ Strategic Planning: Schedule events, renovations, and adjust pricing or availability based on data.

- Pricing Strategies:

➔ Guest Type-Based Pricing: Rates vary by guest preferences, purpose, demographics, etc.
➔ Occupancy-Based Pricing: Higher rates in high-demand periods, lower in low seasons.
➔ Segment Strategy: Different rates for the same room based on guest category (e.g., corporate discounts).
➔ Dynamic Pricing: Real-time price adjustment based on demand, competition, and external factors.
➔ Cancellation Policy Pricing: Reduced rates for non-refundable bookings, mitigating cancellation losses.

Limitation:

This analysis, based on the 'hotel booking' dataset, presents several limitations that should be considered. The dataset, while cleaned and standardized, may still contain errors or biases. The findings are specific to the dataset's time frame and may not account for evolving trends. The scope of the analysis is focused on select aspects of hotel management, excluding external factors like economic conditions or regulatory changes. The predictive model employed has its constraints, including linearity assumptions and reliance on historical data.

Additionally, the analysis does not consider real-time external events that may impact hotel performance. Generalizability of the findings should be exercised with caution, and ethical data handling and privacy considerations are assumed. Temporal relevance is a factor, given the historical nature of the data. Future research can address these limitations for a more comprehensive understanding of hotel management.

Conclusion:

In this data-driven exploration of hotel management using the 'hotel booking' dataset, we have uncovered valuable insights that can guide hoteliers in optimizing their operations and enhancing guest satisfaction. Through thorough analysis, we have addressed key aspects of hotel management, including customer demographics, seasonal trends, customer segmentation, revenue management, and predictive analytics. Our findings reveal that understanding customer preferences, adapting pricing strategies to seasonal demand, and personalizing services are essential components of successful hotel management.

Specifically, we have identified the following key takeaways:

- Customer Demographics: Targeted marketing efforts in regions like Portugal, Great Britain, and France, which contribute significantly to guest origins, can be highly effective. Additionally, recognizing the dominance of "Transient" customers (75.1% of guests) suggests the importance of personalization and tailored experiences.
- Seasonal Trends: Peak booking periods during summer months call for optimized pricing strategies, while the prevalence of last-minute bookings underscores the need for agility and attractive last-minute deals.
- Customer Segmentation: Recognizing that "Transient" customers exhibit the highest cancellation rates (approximately 41%) necessitates a focused approach to manage cancellations within this segment. Encouraging guests to make special requests can significantly reduce cancellation rates and enhance guest loyalty.

- Revenue Management: Fluctuating Average Daily Rates (ADR) throughout the year highlight the importance of dynamic pricing strategies to maximize revenue. It is essential to recognize that "Transient" customers contribute significantly more to revenue than other customer types.
- Predictive Analytics: The logistic regression model, with an accuracy of 79% and a strong ability to predict non-cancellations, provides a valuable

tool for forecasting and managing cancellations. Key predictors of cancellations include lead time, previous cancellations, and the absence of special requests.

In conclusion, this analysis underscores the power of data-driven decision-making in the hospitality industry. By leveraging these insights, hotel management can tailor their strategies, offering personalized experiences, optimizing pricing, and enhancing service quality to meet guest expectations. The ultimate goal is to elevate guest satisfaction, maximize revenue, and maintain a competitive edge in the ever-evolving hospitality landscape.
As hoteliers implement the recommendations derived from this analysis, they will be better positioned to adapt to changing market conditions, optimize occupancy rates, and deliver exceptional guest experiences, ultimately driving success in the hotel industry.

References:

Kotler, P., Bowen, J. T., & Makens, J. C. (2013). Marketing for Hospitality and Tourism. Pearson.
https://www.pearsonhighered.com/assets/preface/0/1/3/5/0135209846.pdf

Cloudbeds. How to use hotel data analytics to improve your property's performance
https://www.cloudbeds.com/articles/hotel-data-analytics/

Jordan Hollander. (2023). Hotel Data Analytics: What You Need to Know About Big Data in Hospitality
https://hoteltechreport.com/news/hotel-data-analytics

Research Gate
13226 PDFs | Review articles in HOSPITALITY INDUSTRY (researchgate.net)

Atlan. (2023). 7 Use Cases of Data Analytics in Hospitality Industry
https://atlan.com/data-analytics-in-hospitality-industry/

https://str.com/data-insights-blog/what-is-revenue-management