-: Hotel Booking Analysis :-

MIS 776 Data Analytics Teaching Project

Group 3

Professor : Richard Young

Presented by: Kajal Dhimmar & Claire Gong

UNLV

Unlocking Hospitality Insights: A Data-Driven Exploration

Executive Summary:

Venturing into the world of hotel management, we are compelled by the promise of data-driven discoveries. Our focus is the 'hotel booking"' dataset from Kaggle, a treasure trove of information containing 119,390 entries and 31 variables for City and Resort hotels. Our aim is clear: to employ data analysis and machine learning to uncover valuable patterns, trends, and strategies within this dataset. From deciphering customer preferences to optimizing revenue, our mission is to extract practical insights that can enhance hotel management practices and make a meaningful impact on the broader hospitality industry. Join us on this journey as we explore the hidden potential within hotel booking data, with the belief that data-driven decisions can elevate guest experiences and drive success in the hotel industry.

Problem Statement:

How can a hotel leverage data analytics to optimize booking management, reduce cancellation rates, and maximize revenue while enhancing guest satisfaction?

Developing a comprehensive analytical model using the 'hotel booking' dataset to predict the likelihood of booking cancellations and optimize room pricing strategies. The model will integrate predictive analytics to forecast cancellations based on customer profiles, booking patterns, and seasonal trends. Additionally, the analysis will include customer segmentation to identify distinct customer groups and their preferences. The final objective is to leverage these insights to implement dynamic pricing strategies that maximize revenue while maintaining high occupancy rates. This model will aid in enhancing the hotel's revenue management system and provide a competitive edge in the market.

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:

Our analysis commenced with the loading of the 'hotel_booking.csv' dataset using Pandas. After a brief examination of column names and types, we addressed missing values. The 'agent' and 'company' columns, plagued by extensive gaps, were removed. Missing country data was standardized to 'Unknown,' and for 'children,' missing values were replaced with the mode. These measures ensured dataset integrity, paving the way for our subsequent analysis of hotel reservation insights.

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.

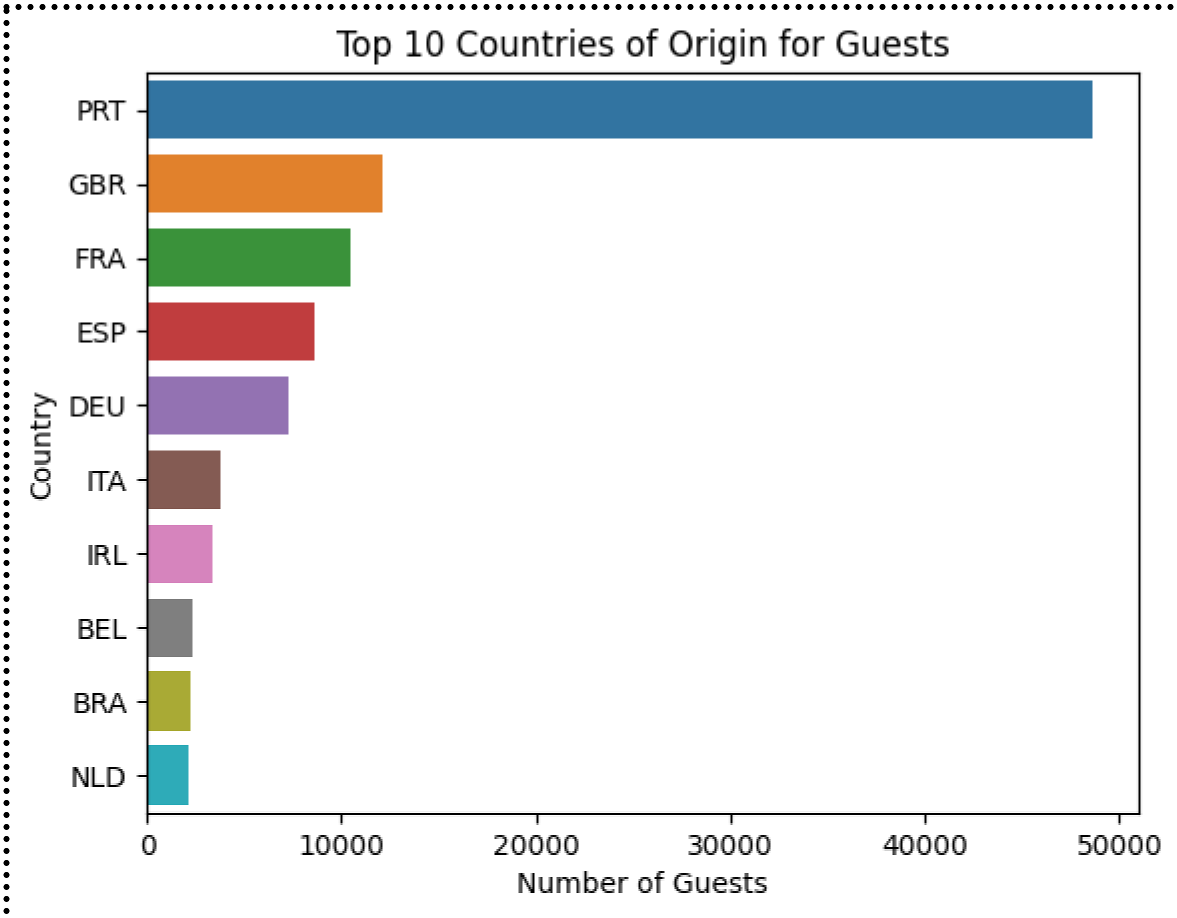

Figure 1.1.1_Geographical Distribution of Guests

Customer Types

Moving to customer types, as depicted in both the bar and pie charts, 'Transient' customers form the largest segment, accounting for a staggering 75.1% of all customers. 'Transient-Party' comes next with 21%, followed by 'Contract' and 'Group' customers with 3.4% and 0.5%, respectively. This indicates a predominantly individual or small-group travel pattern among our clientele, guiding us to potentially focus on personalization and individualized guest experiences.

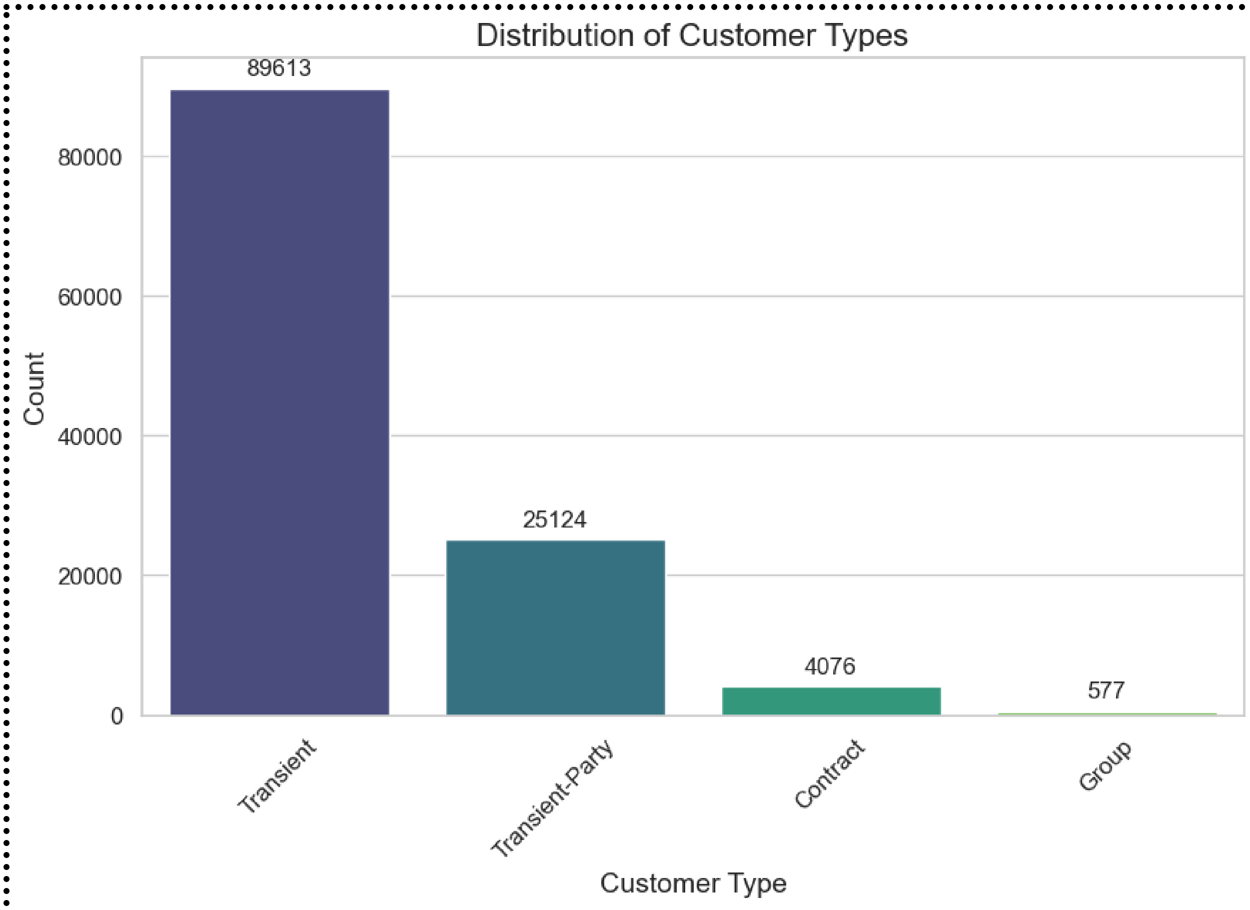

Figure 1.2.1_Distribution of Customer_types

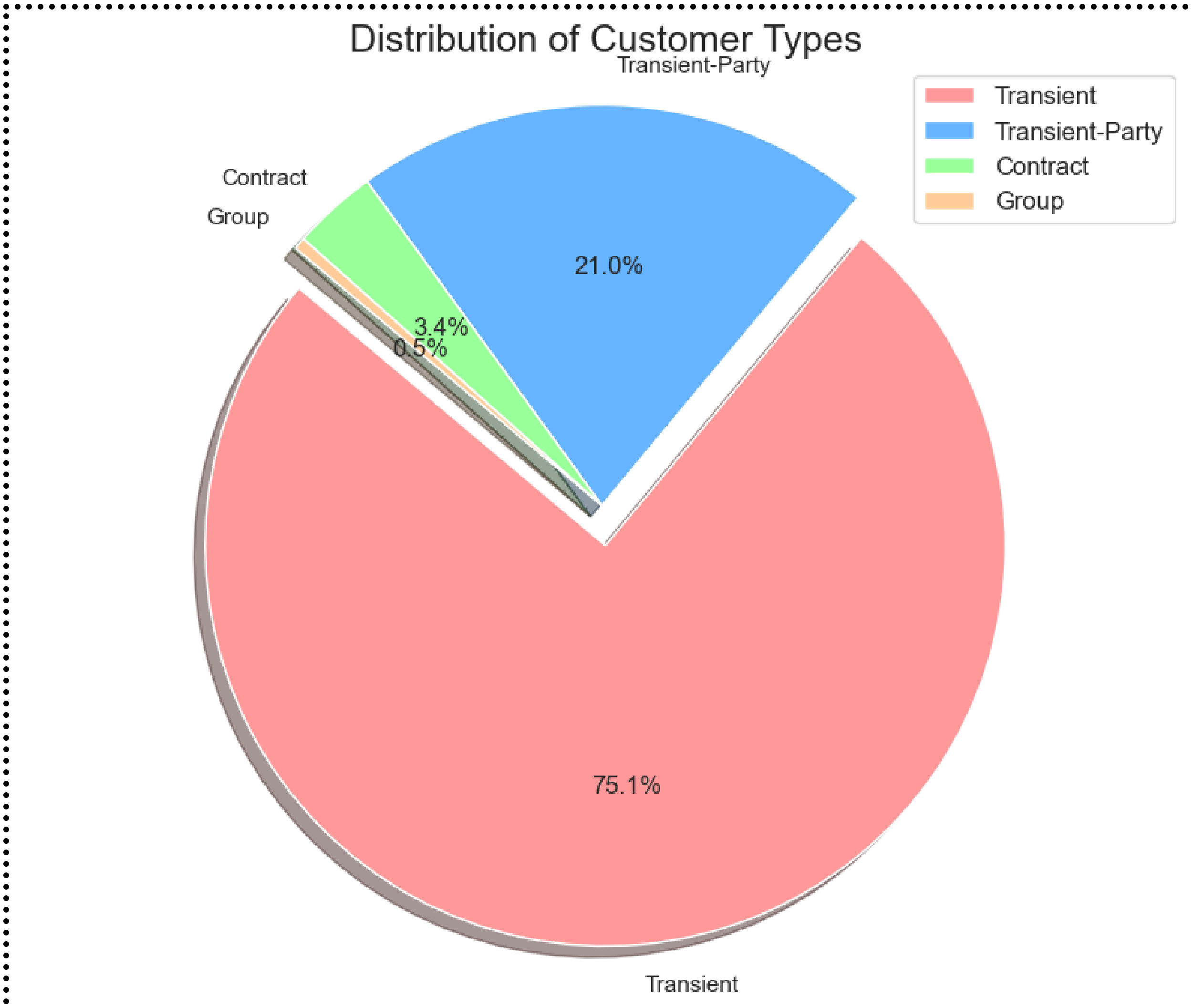

Figure 1.2.2_Distribution of Customer_types

Room Preferences and Special Requests

The data on room preferences and special requests presents a clear picture of guest priorities. Our analysis of room type preferences shows a pronounced trend towards a specific room type, which emerges as the most commonly booked. Ensuring the availability and maintaining the high standards of these preferred rooms could be a key factor in boosting guest satisfaction and encouraging repeat visits. Simultaneously, the special requests data reveals that a significant number of guests opt not to request additional services. This finding suggests an untapped opportunity to promote these offerings more actively, potentially enhancing the guest experience and fostering loyalty. By highlighting the possibility of tailoring their stay through special requests, we can not only personalize our service but also incrementally increase revenue.

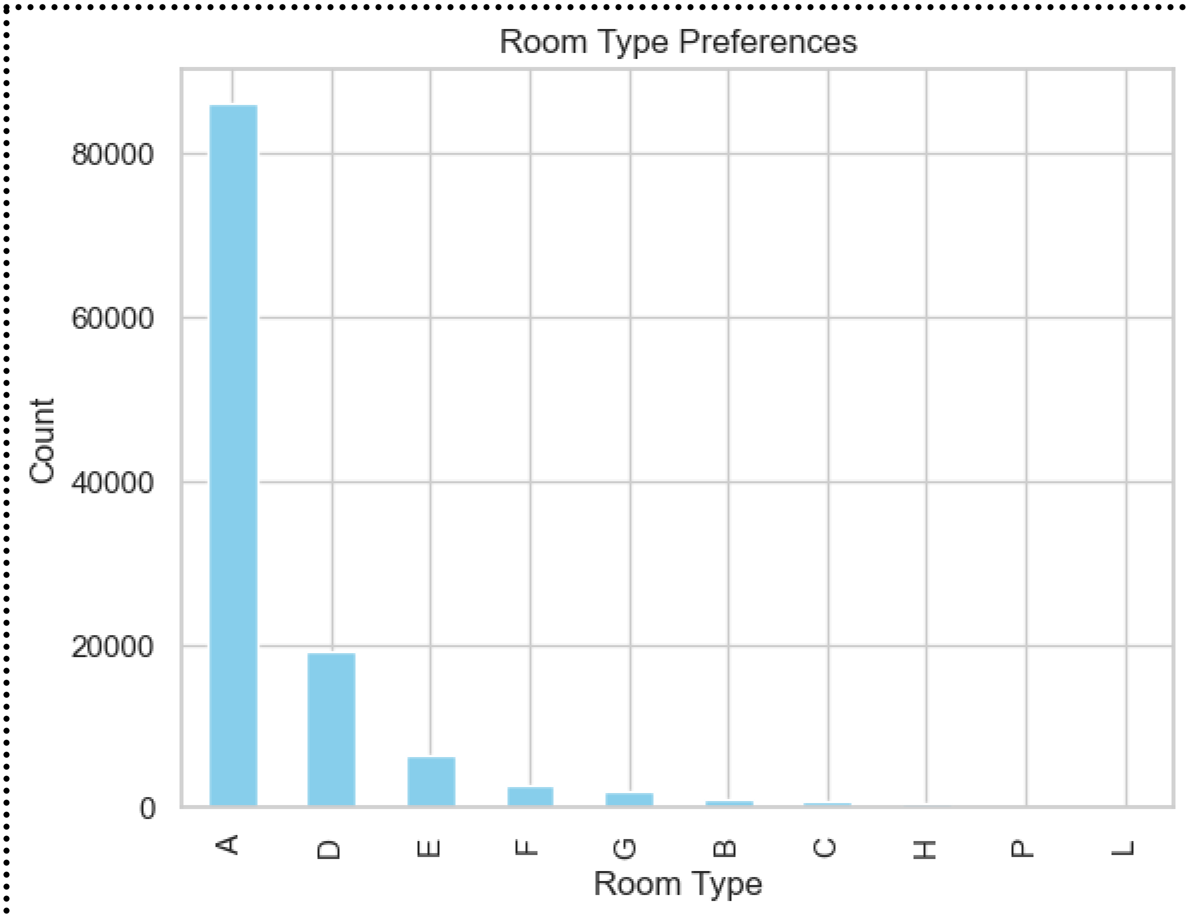

Figure 1.3.1_Room_Type_Preference

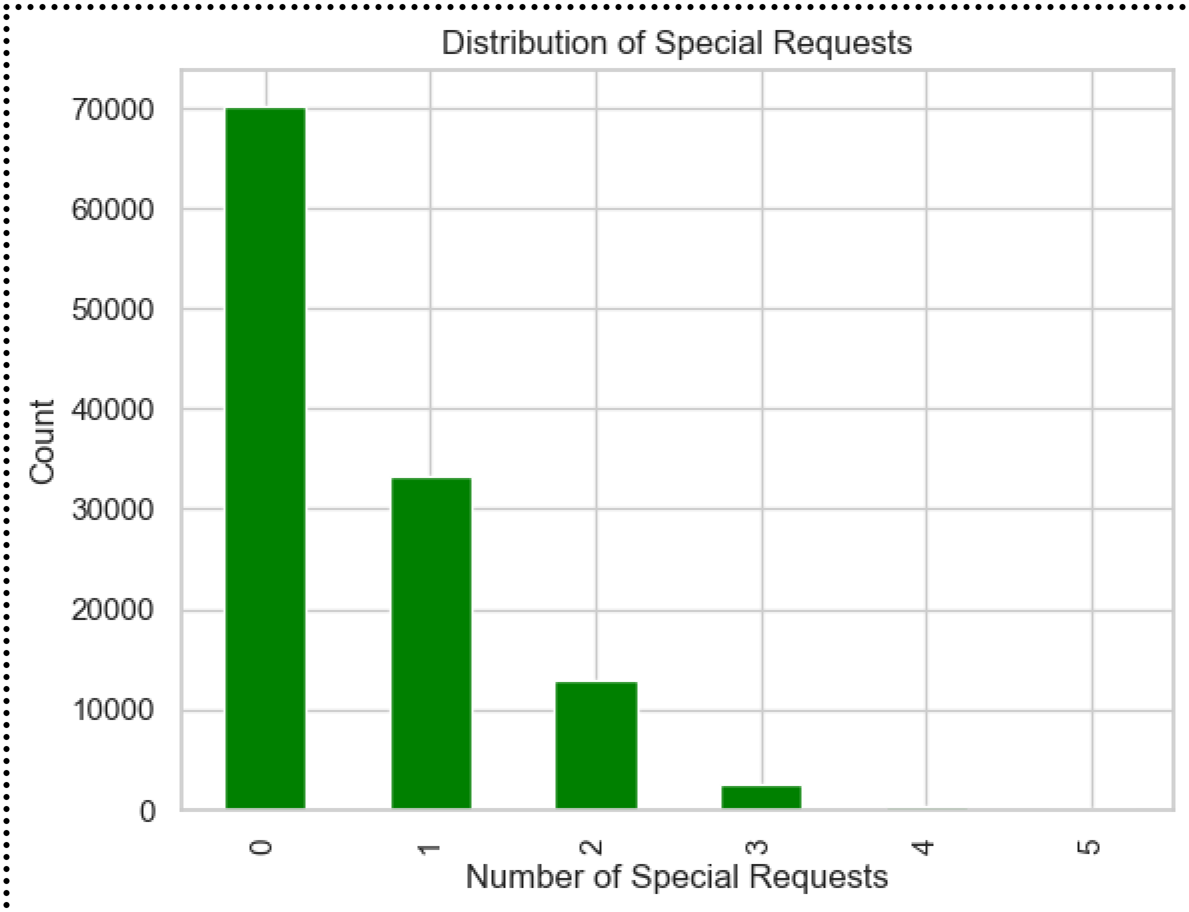

Figure 1.3.2_Distribution of Special_Requests

2. **Seasonal Trend Analysis:**

Our evaluation of booking dates and lead times has uncovered distinct patterns, crucial for developing effective pricing and promotional strategies.

Peak and Off-Peak Seasons

The "Number of Bookings per Month" bar chart illustrates the ebb and flow of guest bookings throughout the year. Peak booking periods are evident in the months of August, July, and May, with the highest number occurring in August. This suggests a high demand for accommodations during these summer months, which could be attributed to the summer holiday season when tourists are more likely to travel. Conversely, the months showing fewer bookings, such as January, December, and November, represent off-peak periods where the hotel could potentially implement promotional campaigns to increase occupancy rates.

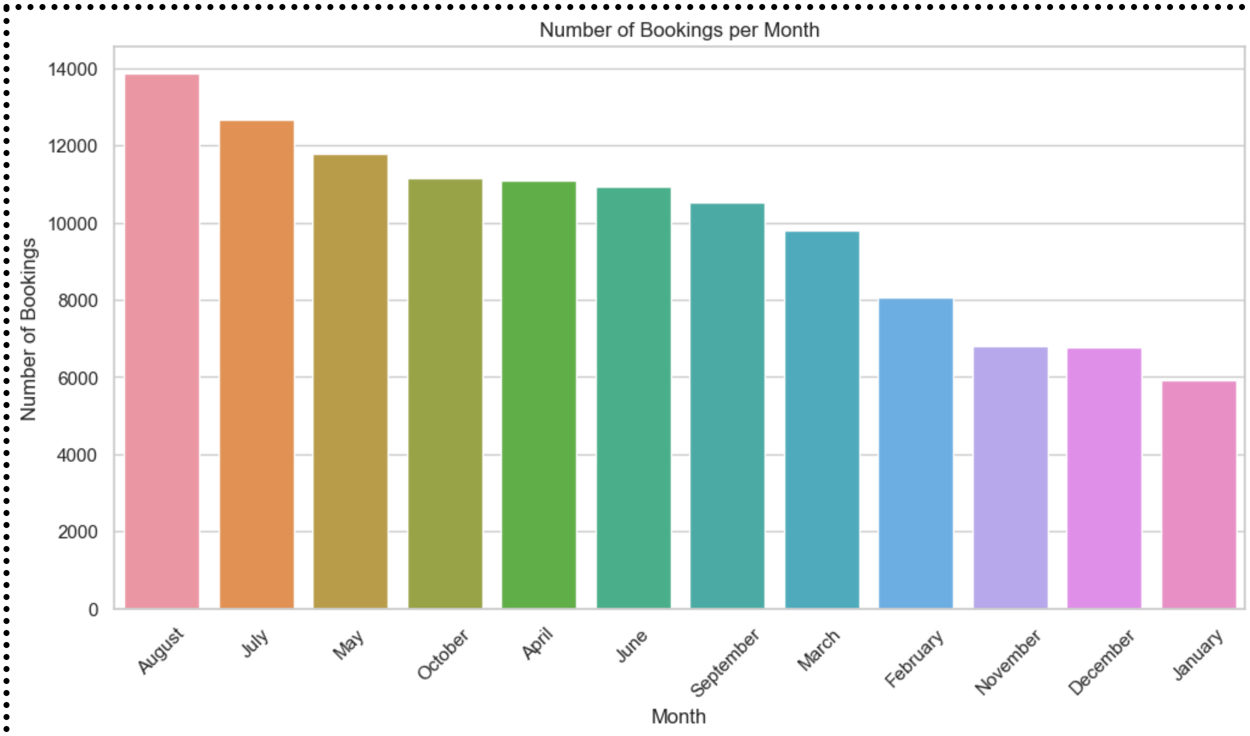

Figure 2.1.1_Number_of_Bookingd_per_Month

Booking Lead Times

The "Distribution of Lead Time for Bookings" histogram indicates that a significant number of bookings are made with a relatively short lead time, peaking at 0-10 days. This finding highlights the importance of last-minute booking strategies and the potential to capture this market with timely offers. However, the histogram also shows a tail of bookings extending towards longer lead times, demonstrating the necessity for early-bird promotions targeting planners who book their stays well in advance.

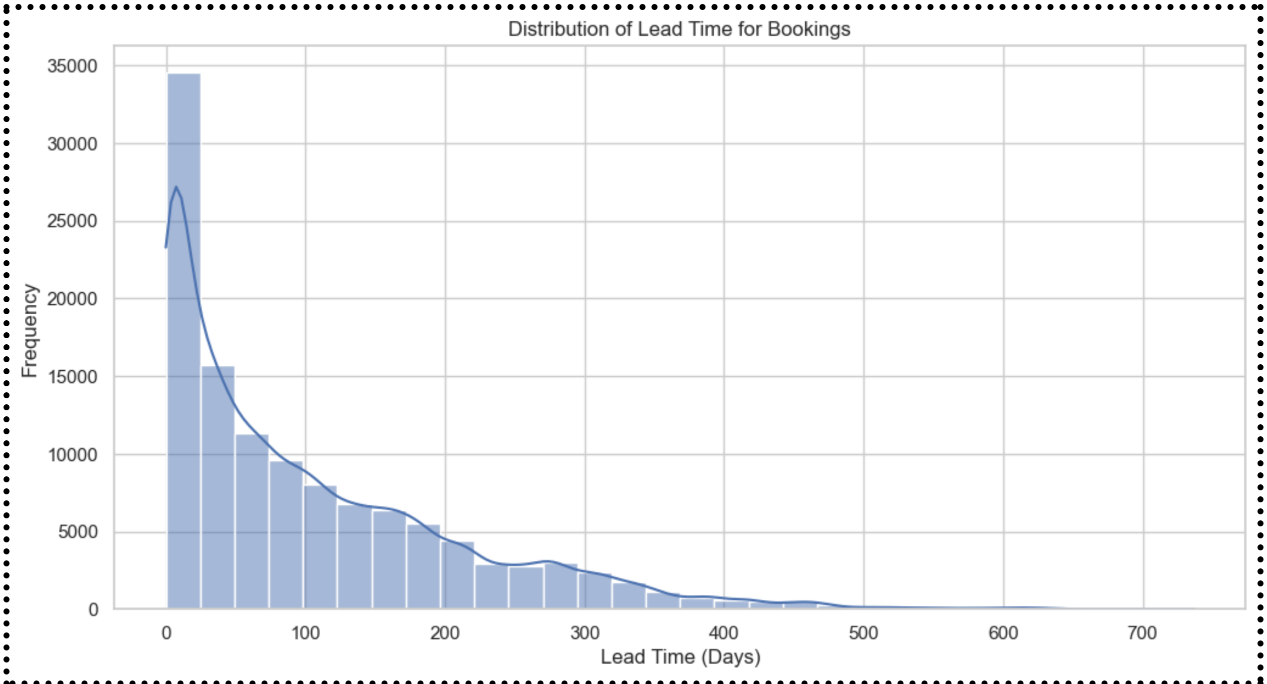

Figure 2.2.1_Distribution_of_Lead_Time_for_Bookings

Yearly Booking Patterns

The heatmap titled "Monthly Bookings Heatmap" provides a year-over-year comparison, revealing both consistent trends and anomalies. For instance, there is a notable increase in bookings from June to October for each year shown, with the highest concentrations observed in October 2016 and May 2016, suggesting peak demand during these periods. The heatmap also shows a stark absence of data for the months of January to June in 2015 and for September to December in 2017, which could be due to the dataset not covering these months or the hotel being closed or not accepting bookings during those periods. This comparative view enables us to not only anticipate demand but also to identify unusual fluctuations that could inform our marketing and operational preparedness.

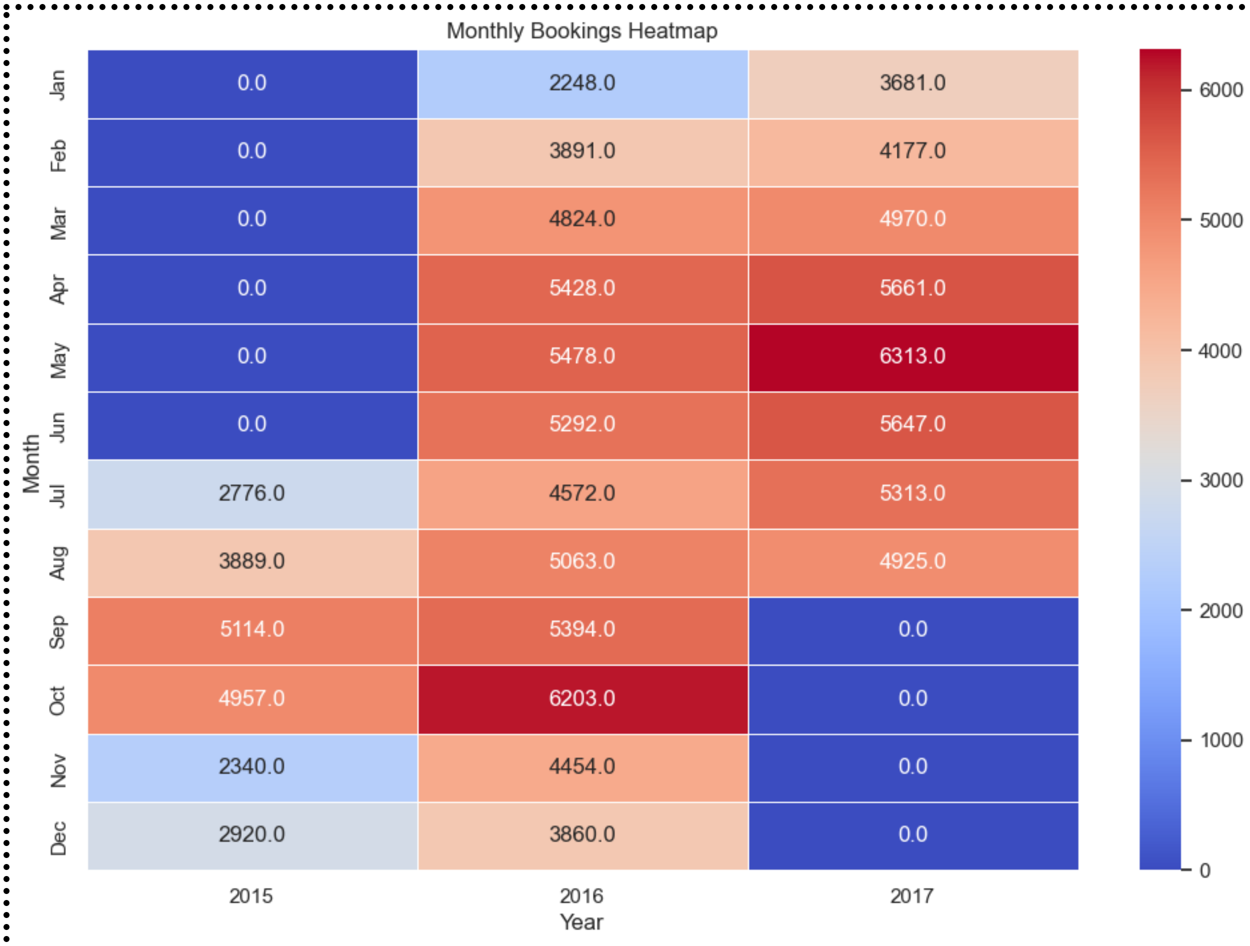

Figure 2.3.1_Monthly_Booking_Heat_Map

3. Customer Segmentation Cancellation Analytics:

a. Transient customers means the customer who are temporarily traveling have the highest cancellation rate almost ~ 41 %

b. The more special requests a booking has, the lower the cancellation rate. Bookings with no special requests have a cancellation rate of about 47.72%, while those with 5 special requests have a rate of just 5.00%.

c. Also the number of bookings and cancellation rates are highly correlated. The higher number of bookings tends to have a higher number of cancellation rates and vice versa.

Basic Customer Segmentation Analysis

Looking at the segmentation by country, Portuguese guests lead in the number of bookings, but they also have the highest cancellation rate among the top five countries. This could indicate 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):

country	Number_of_Bookings	Average_Stay_Days	Cancellation_Rate
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

Figure 3.1.1_Segmentation_by_country

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.

```
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
```

Figure 3.1.2_Segmentation_by_customer_type

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.

```
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
```

Figure 3.1.3_Monthly_Booking_Heat_Map

Lead Time Analysis

The plot shows the relationship between lead time and cancellation rate. A general trend can be observed where longer lead times are associated with higher cancellation rates.
These insights can be very useful for understanding the behavior of different customer segments and how various factors influence cancellation rates.We can use this information for targeted strategies to reduce cancellations or for improving booking policies.

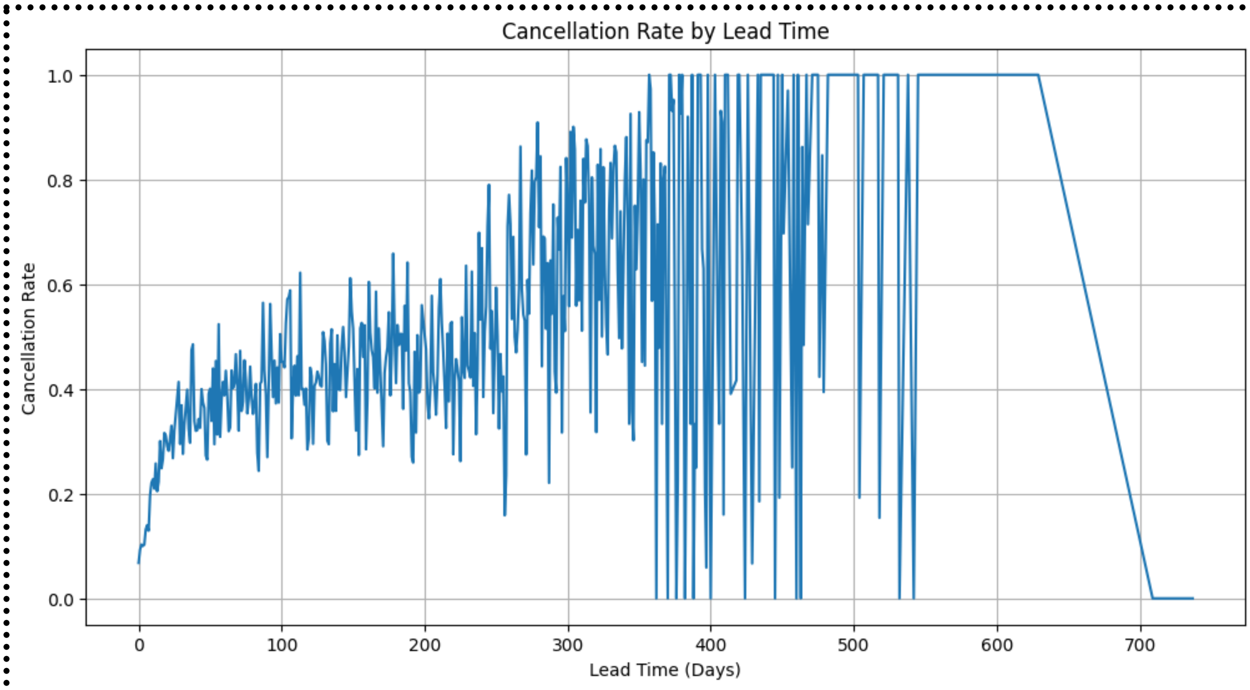

Figure 3.2.1_cancellation_rate_by_lead_time

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.

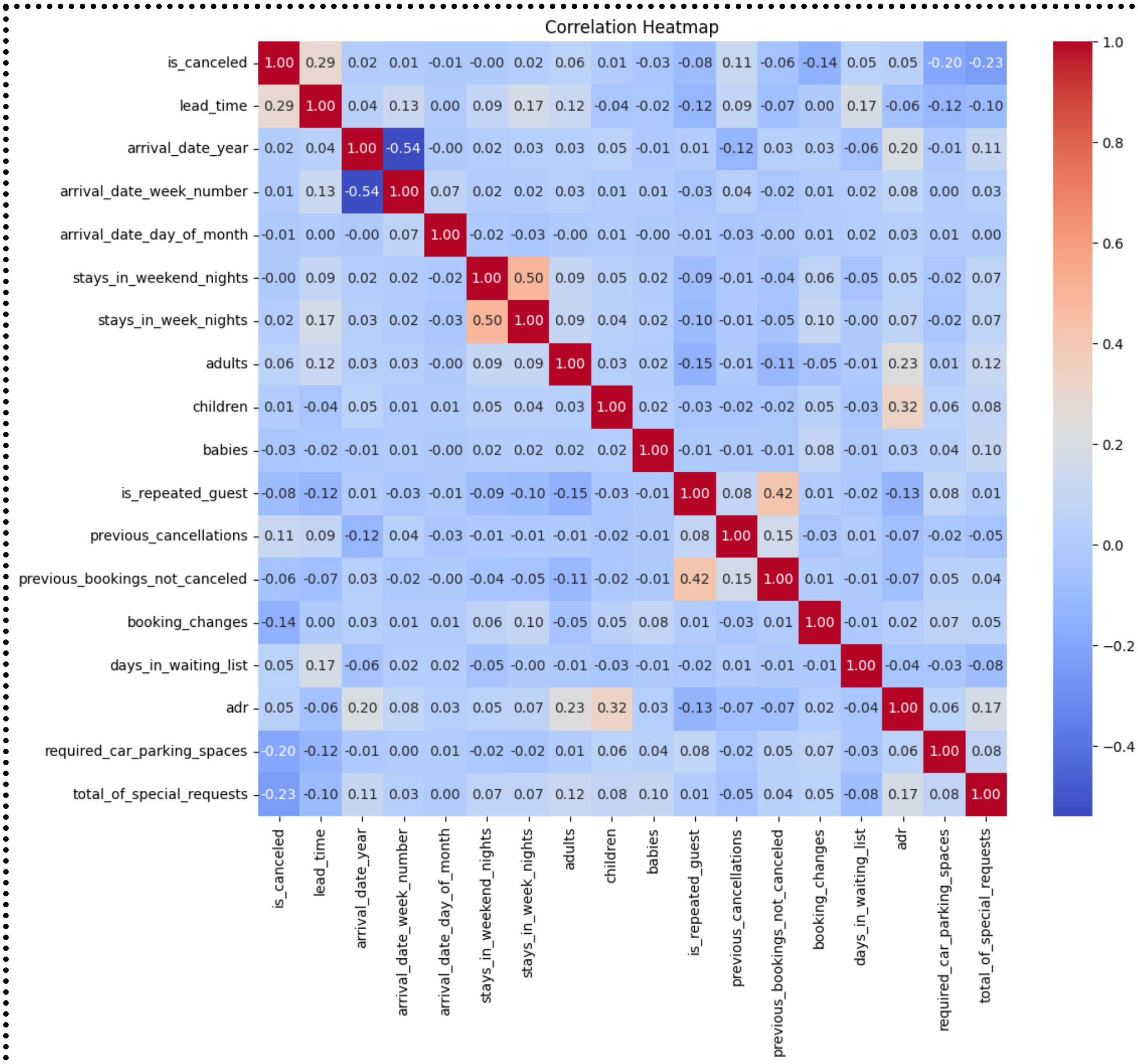

Figure 3.3.1_Correlation_Heatmap

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.

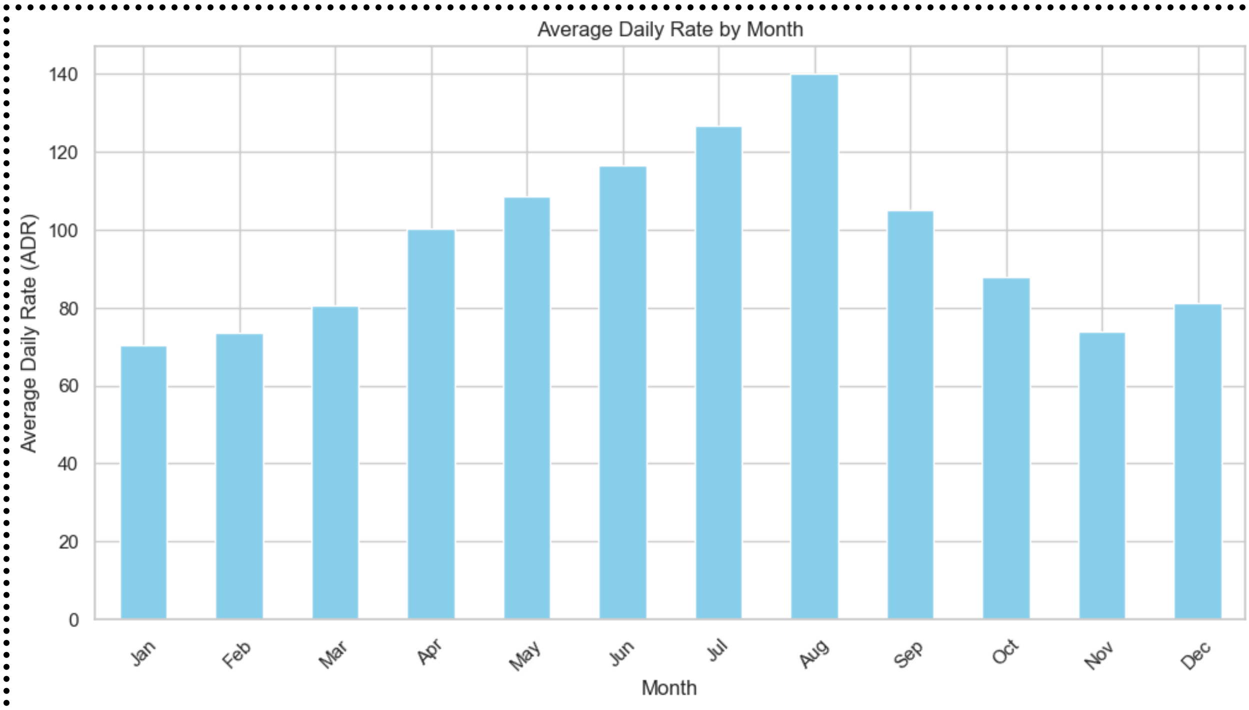

Figure 4.1.1_ADR_by_Month

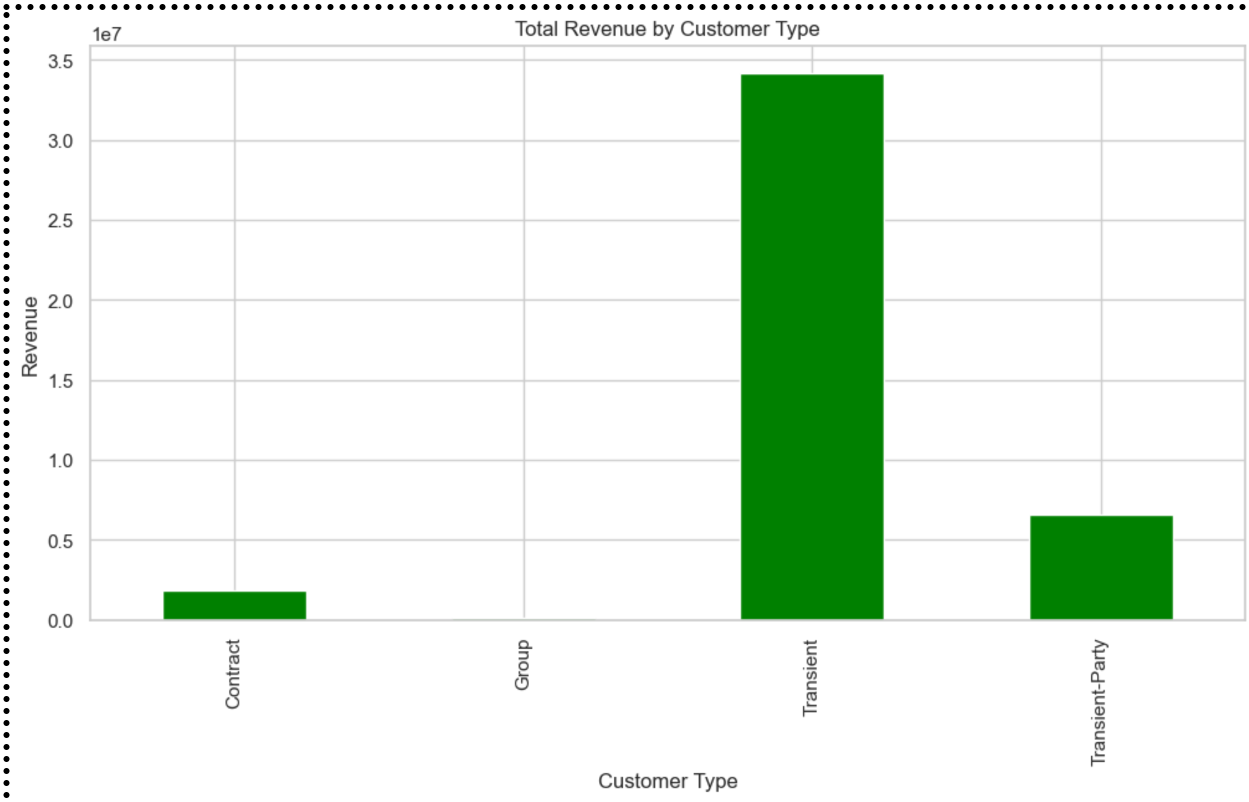

Figure 4.1.2_Total_revenue_by_customer_Type

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): 4,752 were wrongly predicted as non-canceled (false positives), and 6,376 were correctly predicted as canceled (true negatives).

ROC AUC Score:

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     14907
           1       0.81      0.58      0.67      8971

    accuracy                           0.79     23878
   macro avg       0.80      0.75      0.76     23878
weighted avg       0.79      0.79      0.78     23878

Confusion Matrix:
 [[13716  1191]
 [ 3804  5167]]
ROC AUC Score: 0.8534449349351705
```

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

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

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