

# Hotel Booking Trends Analysis and Optimization Strategies

## - A Project Report

The hospitality industry is constantly evolving, and for hotel chains, understanding booking patterns is crucial to improving operations and maximizing revenue. This exploratory data analysis (EDA) focuses on uncovering insights from a dataset related to hotel bookings, aiming to address the challenges faced by the hotel management team. The dataset includes vital information such as booking status, lead time, customer demographics, market segments, arrival dates, and cancellation details.

The goal of this project is to leverage this data to gain a deeper understanding of booking trends and behaviors. Specifically, this analysis will explore factors influencing booking lead time, early booking patterns, cancellation rates, and occupancy trends. Identifying these patterns will help to provide actionable recommendations to help the hotel management reduce cancellations, increase early bookings, optimize pricing strategies, and ultimately improve occupancy rates. The insights drawn will also support data-driven decision-making, enabling the hotel to better predict demand and allocate resources more efficiently.

### Data Overview

The dataset contains 32 columns and 119,390 rows, with a mix of numeric (e.g., `average_daily_rate`) and categorical (e.g., `arrival_date_month`, `hotel_type`) data types. Below is a summary of the most important columns:

- **Hotel Information:** The dataset differentiates between two types of hotels (Resort Hotel and City Hotel).
- **Booking Status:** The variable `is_canceled` indicates whether a booking was canceled (1) or not (0), while `reservation_status` tracks the final status of the reservation (e.g., Canceled, Check-Out, No-Show).

- **Booking Dates:** Includes detailed date-related information such as **lead\_time** (days from booking to arrival), **arrival\_date\_year**, **arrival\_date\_month**, and the specific day of arrival (**arrival\_date\_day\_of\_month**).
- **Guest Demographics:** Details about the guests include the number of **adults**, **children**, and **babies**, as well as whether the guest is a **repeated\_guest**.
- **Stay Details:** Captures the number of weekends (**stays\_in\_weekend\_nights**) and week nights (**stays\_in\_week\_nights**) a guest booked to stay, as well as any special requests (**total\_of\_special\_requests**).
- **Room and Meal Information:** Includes details on the reserved and assigned room types (**reserved\_room\_type**, **assigned\_room\_type**) and meal preferences (**meal**), with various meal package options.
- **Market Segment and Distribution Channel:** The **market\_segment** and **distribution\_channel** variables identify how the booking was made (e.g., Travel Agents, Tour Operators).
- **Booking and Cancellation History:** Includes data on previous cancellations (**previous\_cancellations**) and bookings that were not canceled (**previous\_bookings\_not\_canceled**).
- **Additional Booking Characteristics:** Information such as deposit type (**deposit\_type**), whether a deposit was made to guarantee the booking, **days\_in\_waiting\_list**, and whether the guest required car parking spaces (**required\_car\_parking\_spaces**) is also provided.

## Data Cleaning and Preprocessing

To ensure the dataset is ready for analysis, several data cleaning and preprocessing steps were performed as follows:

### General Cleaning

- Ensured all columns had the correct and consistent data types.
- Fixed spelling errors in **meal**, **market\_segment**, and **distribution\_channel** to standardize inputs.
- Trimmed whitespace across the dataset using the Google Sheets' Data cleanup tool. Removed extra spaces within inputs manually.

### Standardization

- Verified email standardization using regular expressions. Converted all emails to lowercase for consistency.
- Replaced ISO country codes with full country names.

### New Columns

- Added a season column categorizing dates into summer, fall, winter, or spring.
- Created an arrival\_date column by combining arrival\_date\_year, arrival\_date\_month, and arrival\_date\_day\_of\_month.

### Removed Columns

- The arrival\_date\_year, arrival\_date\_month, and arrival\_date\_day\_of\_month columns were removed after arrival\_date column creation.

### Handling Missing Data

- Children column: Imputed missing values with 0, the most common entry.
- Name column: Extracted missing names from the corresponding email.
- Country column: Replaced missing values with "Unknown".
- Stayed\_in\_weekend\_nights: Missing values replaced with 0.
- Agent and Company columns: Filled missing values with "Null".

### Identified Issues

- Agent and Company columns showed inconsistencies.
  - When an agent ID was present, the company ID was often missing, and vice versa.
  - A significant number of rows had neither ID, suggesting a possible data entry oversight

## Exploratory Data Analysis With SQL

1. Which market segment generates the highest revenue?
  - a. **Online Travel Agencies (Online TA)** dominate the market with **\$23,941,251** in revenue, which is a significant share compared to other segments.
  - b. **Complementary** services bring in **\$5,083**, making it the smallest revenue generator.

Market Segment	Booking Revenue
Online TA	\$23,941,251
Offline TA/TO	\$8,146,333
Direct	\$5,093,028
Groups	\$4,669,764
Corporate	\$774,295
Aviation	\$87,446
Complementary	\$5,083

### Insight:

- a. Online channels represent the largest share of revenue. This suggests that e-commerce and digital booking platforms have a substantial influence on the hotel business, likely reflecting broader digital consumer trends.
  - b. Aviation and complementary segments are relatively minor.
    - i. This indicates that the core of the business might be centered around broader travel services rather than flight-related or additional services.
    - ii. Also, it may suggest complementary services are less central to the business model or may be supplementary rather than a primary offering.
- 
2. How many bookings were made per market segment?

Market Segment	Total Bookings
Online TA	56478
Offline TA/TO	24217
Groups	19811
Direct	12606
Corporate	5295
Complementary	743
Aviation	237

### Insights:

- a. The data indicates that consumers prefer booking online, maybe due to convenience, ease of comparison, or possibly the availability of deals and offers that are exclusive to digital platforms.
  - i. The dominance in bookings aligns with the revenue data, meaning investments in improving the online experience. They include user interface, mobile apps, personalized recommendations. This could further boost bookings.
  - ii. Expanding digital marketing campaigns could also capture untapped regional markets.
- b. Even though Complementary services had more bookings than the Aviation segment, it still had significantly less revenue in comparison.
  - i. The higher number of bookings in complementary services indicates strong demand, but with less consistency in value per customer.
  - ii. In contrast, aviation bookings are fewer but cater to a more exclusive market, offering a higher value per booking due to the premium nature of the services.

### 3. What is the distribution of bookings across customer types?

Customer Type	Total Bookings
Transient	89611
Contract	4075
Transient-Party	25124
Group	577

The **Transient** customer type represents individual travelers or those booking one-off trips.

The **Transient-Party** group represents bookings for multiple individuals, such as friends or families traveling together.

**Contract** customers refers to those who have an ongoing agreement or frequent travel needs, such as businesses or organizations booking travel for employees.

The **Group** customer type represents larger travel parties such as tours or corporate retreats.

### Insights:

- a. As the Transient group makes up the bulk of bookings, it's essential to focus on:
  - i. Optimizing the individual travel experience by offering tailored packages, loyalty programs, and personalized services to attract repeat business.

- ii. Improving the booking process to make it seamless can help convert more transient customers.
- b. To improve Transient-Party bookings,
  - i. Offering group-specific packages such as family vacations, multi-room bookings, or group discounts can increase customers.
  - ii. Tailoring marketing efforts towards parties (e.g., promotions or incentives for group bookings) can lead to higher conversion rates.
  - iii. This group may also benefit from services like shared accommodation.

4. What are the monthly trends in booking numbers?

Year	Month	Total Bookings
2015	July	2775
2015	Aug	3889
2015	Sept	5114
2015	Oct	4957
2015	Nov	2340
2015	Dec	2919
2016	Jan	2248
2016	Feb	3891
2016	Mar	4823
2016	Apr	5428
2016	May	5478
2016	June	5292
2016	July	4572
2016	Aug	5063
2016	Sept	5394
2016	Oct	6203
2016	Nov	4454
2016	Dec	3860
2017	Jan	3681
2017	Feb	4177
2017	Mar	4970
2017	Apr	5661
2017	May	6313
2017	June	5647
2017	July	5313
2017	Aug	4925

**Insights:**

- a. The monthly booking trends show clear seasonal variations, with peak demand occurring in April, May, and June, indicating a high season for hotel bookings. In contrast, a notable decline is observed in November, December, and January, suggesting a low season.
- b. The pattern appears cyclical, with similar high and low booking periods repeating each year.
  - i. The mid-year months consistently attract the highest number of bookings, while the winter months experience lower volumes. This indicates predictable fluctuations in demand.
- c. Hotels should adjust pricing and marketing strategies based on demand fluctuations.
  - i. Peak months may require increased staffing and resource allocation to accommodate higher bookings.
  - ii. Low-demand periods present an opportunity for promotional offers to attract more guests and maintain steady occupancy rates.

## Customer Behaviour Analysis

5. What is the average lead time for bookings?
  - a. The average lead time for bookings was 104 days.
  - b. An analysis of the average lead time for bookings per market segment revealed the following:

Market Segment	Avg Lead Time (days)
Groups	187
Offline TA/TO	135
Online TA	83
Direct	50
Corporate	22
Complementary	13
Aviation	4

**Insights:**

- a. Since group and offline TA/TO bookings happen well in advance, travel businesses should focus on early booking promotions, group discounts, and customized travel planning well before the travel dates to maximize value.
- b. For market segments with shorter lead times, travel companies can focus on last-minute deals and dynamic pricing. They should also ensure their systems are equipped to handle spontaneous bookings efficiently, providing the flexibility and options these travelers expect

6. Which room types have the highest cancellation rate?

Room Type	Cancellations	Cancellation Rate
A	33628	76.044
D	6102	13.799
E	1914	4.328
F	880	1.990
G	763	1.725
B	368	0.832
C	308	0.696
H	245	0.554
P	12	0.027
L	2	0.005

- a. **Room Type A** has the highest cancellation rate at **76.04%**.
- b. This is followed by Room Type D with a cancellation rate of 13.80%, and Room Type E with a cancellation rate of 4.33%.
- c. The cancellation rate is directly proportional to the number of bookings of the room type.

**Insights:**

- a. Room A stands out with an exceptionally high cancellation rate, which could indicate issues with customer commitment or pricing.
  - i. Implementing more restrictive cancellation policies for this room type or analyzing why customers are canceling could provide valuable insights.



- b. Room types E, G, and F show lower cancellations, suggesting they are more stable options. These room types could represent an opportunity for targeted upselling or longer stays.
- c. Room types C, B, and H show the lowest cancellation rates, indicating a premium clientele or non-refundable booking policies.
  - i. Marketing these rooms as exclusive or luxury options could further capitalize on their strong booking stability.

7. Which countries have the most bookings?

Country	Total Booking
Portugal	48588
United Kingdom	12129
France	10415
Spain	8568
Germany	7287
Italy	3766
Ireland	3375
Belgium	2342
China	2278
Brazil	2224

**Insights:**

- a. Portugal has an overwhelming lead in the number of bookings, with more than 48,000 bookings. This suggests that Portugal is a top tourist destination or has a strong local travel market.
  - i. It's important to maintain and expand marketing efforts in Portugal to continue benefiting from this dominant position.
  - ii. Operations in Portugal should be further strengthened by investing in local promotions, tailored experiences, and partnerships with local hotels, airlines, or tourism agencies.
- b. Britain, France, and Spain have substantial numbers of bookings, indicating that these countries are also strong markets for tourism and travel

- i. These countries represent stable markets, and efforts should be made to increase engagement by focusing on specific tourist attractions like cultural events in France or beach resorts in Spain.
    - ii. Marketing in these regions could be more aggressive in order to outperform competition.
  - c. The remaining countries might have a steady but niche market in terms of tourism and they represent emerging opportunities.
8. What is the ratio of repeat customers to new customers?
- a. Number of Repeat Customers: 3,810  
Number of New Customers: 115,578
  - b. The ratio of repeat customers to new customers is:
 
$$\frac{\text{Repeat Customers}}{\text{New Customers}}$$

$$3,810 / 115,580 = 0.033 \approx 3.3\%$$

**Insights:**

- a. The ratio of 3.3% of repeat customers to new customers suggests that the hotels primarily attract new customers.
  - i. This highlights a need for better retention strategies, as retaining customers is often more cost-effective than acquiring new ones.
- b. Since the proportion of repeat customers is relatively low, there might be an opportunity to implement or improve loyalty programs aimed at converting more new customers into repeat guests.
- c. Exploring the reasons why customers are not returning or identifying ways to improve the customer experience could be valuable next steps.

## Revenue Trend Analysis

9. Which room types generate the highest revenue?

Room Type	Booking Revenue
A	\$22,197,039
D	\$10,334,072
E	\$4,029,306
F	\$2,043,643
G	\$1,773,682
C	\$1,145,688
B	\$598,076
H	\$499,701
I	\$52,401
K	\$43,509
P	\$75
L	\$8

- a. **Room Type A** has the highest **total revenue** at **\$22,197,039**, which is significantly higher than all other room types. This is because it also is the most booked room.
- b. Room Type A also had a low average revenue gained suggesting a wide customer base for its low cost

10. During which seasons is the revenue highest?

- a. There is significant variation in hotel revenue across different seasons. Summer, a peak travel period with high demand generates the highest revenue. Spring and Fall follow with moderate revenue levels, indicating steady but lower bookings compared to summer. Winter records the lowest revenue, likely due to reduced travel activity during colder months.

### Insights:

- a. The hotels should optimize pricing and marketing strategies during summer while introducing promotional offers or discounts in winter to attract more bookings and sustain revenue.

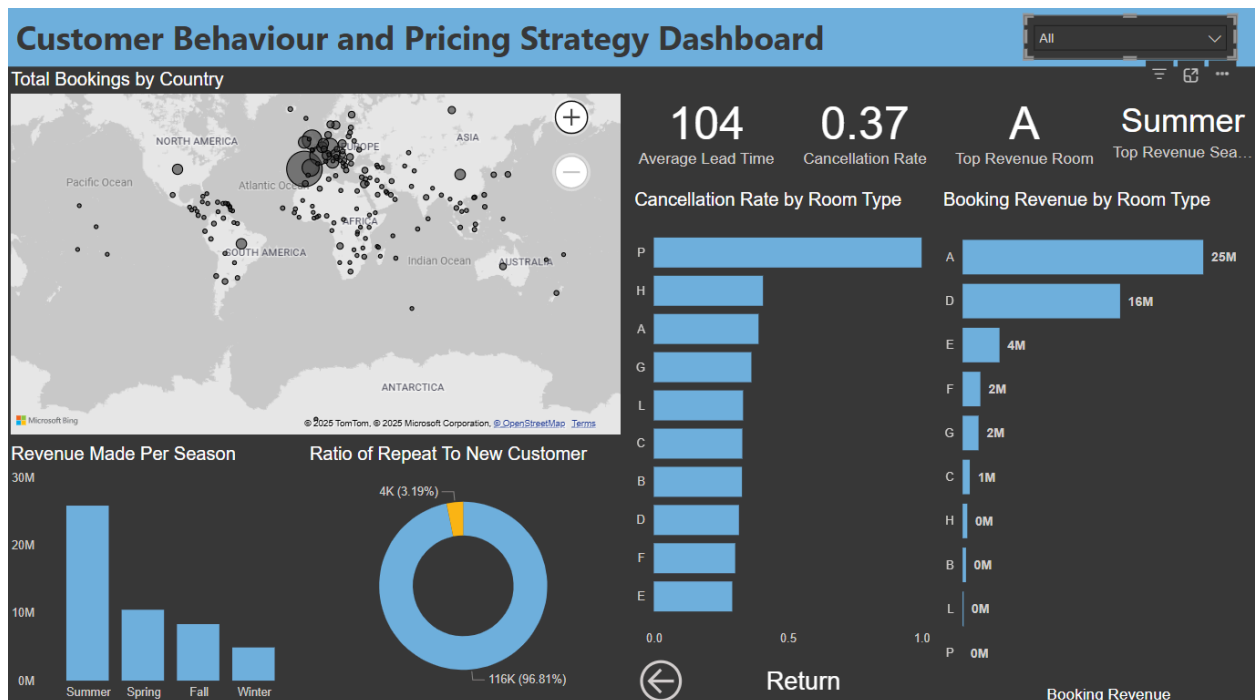
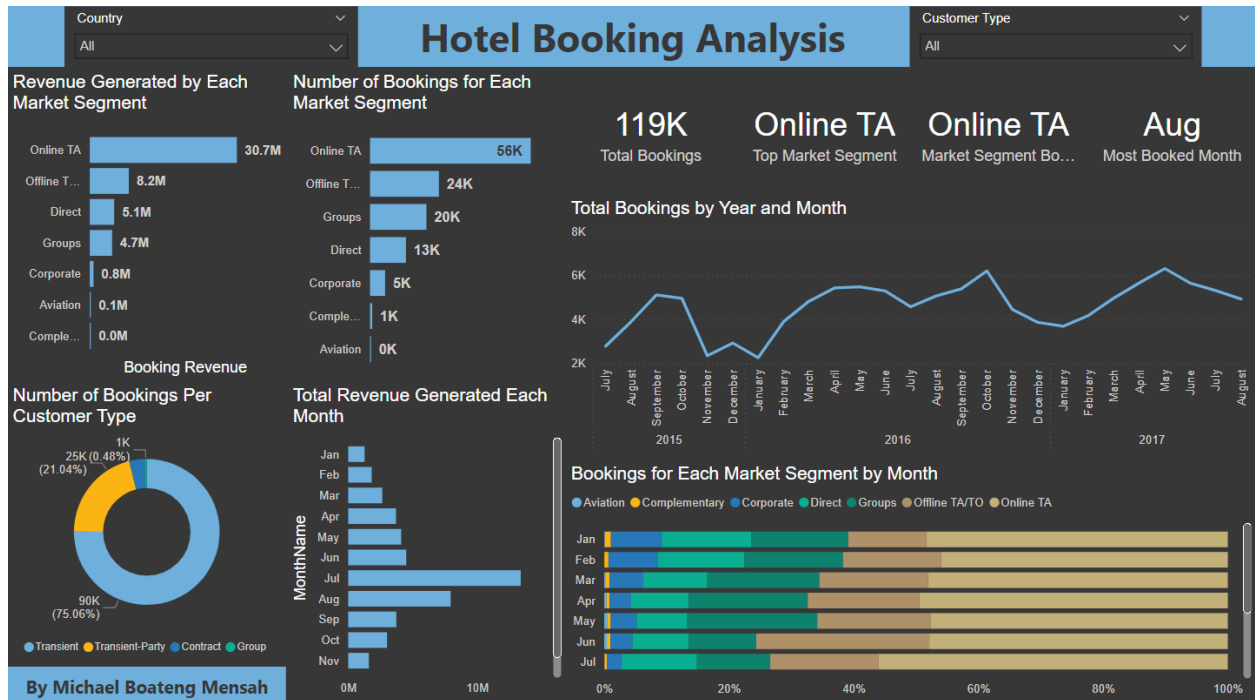
- b. Additionally, staffing and resource allocation should align with these seasonal trends to maximize efficiency and profitability.

## Recommendations

1. For Room Type A:
  - a. Analyze customer feedback for Room Type A to determine if pricing or service improvements can reduce cancellations.
  - b. Implement stricter booking policies (e.g., requiring deposits or non-refundable options).
2. For Reducing Cancellations:
  - a. Offer incentives (such as discounts on future bookings) for non-cancellation commitments.
3. For Increasing Early Bookings:
  - a. Enhance marketing efforts towards corporate and group customers who typically book well in advance.
  - b. Introduce loyalty programs that offer benefits for early bookings, driving higher advance commitments.
  - c. Launch early-bird promotions and group discounts to reward early reservations.
4. For Optimizing Pricing Strategies:
  - a. Implement dynamic pricing models that adjust rates based on seasonal demand and occupancy levels.
  - b. Offer bundled packages or promotions during low-demand periods (winter) to stimulate bookings.
5. To Improve Occupancy Rates:
  - a. Enhance your online booking platform with personalized offers and a seamless user experience.
  - b. Deploy targeted marketing campaigns during off-peak seasons to boost occupancy.
  - c. Combine flexible pricing strategies with value-added packages to attract a broader customer base.

6. To Better Predict Demand And Efficiently Allocate Resources:
- a. Align resource allocation (e.g., staffing levels, marketing spend) with forecasted demand to maximize operational efficiency.

## Dashboard



## Conclusion

This project provided an in-depth analysis of hotel booking patterns. It focused on key factors such as room type demand, cancellation rates, early booking trends, and pricing optimization. Through exploratory data analysis (EDA), we identified that Room Type A is the most frequently booked due to its affordability. However, its revenue potential can be further enhanced through bundled services and cancellation reduction strategies.

Additionally, high cancellation rates present a challenge to revenue stability. To mitigate this, strategies such as stricter cancellation policies, incentives for non-refundable bookings, and improved customer communication were recommended. Encouraging early bookings through discounts, loyalty programs, and targeted promotions can help the hotel maintain a more predictable revenue stream.

By leveraging data-driven insights, the hotel can optimize pricing strategies, improve occupancy rates, and enhance demand forecasting. Implementing these recommendations will not only increase profitability but also improve overall operational efficiency while ensuring sustainable growth in a competitive hospitality market.

## Appendix

1. [Dataset Link](#)
2. [SQL Queries](#)
3. [Power BI Dashboard](#)
4. [Presentation Link](#)