

Exploratory Data Analysis (EDA) of Hotel Booking Dataset

1. Introduction

This report presents the exploratory data analysis (EDA) of a hotel booking dataset. The objective of the analysis is to understand booking patterns, customer behavior, and factors influencing hotel reservations. The insights derived will help hotels optimize operations, improve customer satisfaction, and increase revenue.

2. Dataset Overview

The dataset contains booking information for city and resort hotels, including details about reservation status, customer demographics, booking lead time, stay duration, special requests, and more.

Key Variables:

- **Hotel:** Type of hotel (City Hotel or Resort Hotel)
- **Is Canceled:** Whether the booking was canceled (1) or not (0)
- **Lead Time:** Number of days between booking date and arrival
- **Arrival Date:** Information about the arrival year, month, and week
- **Stays in Week Nights/Weekend Nights:** Duration of stay
- **Adults/Children/Babies:** Guest composition
- **Meal:** Type of meal plan selected
- **Market Segment & Distribution Channel:** How the booking was made
- **Previous Cancellations & Special Requests:** Customer behavior history
- **Assigned Room Type:** Room type assigned compared to the one reserved
- **ADR (Average Daily Rate):** Revenue per occupied room per night
- **Country:** Guest's country of origin

3. Data Cleaning & Preparation

Prior to conducting EDA, extensive data cleaning was performed to ensure data accuracy and consistency. The following steps were taken:

3.1 Handling Missing Values

- Columns with significant missing values (**company**, **agent**, and **country**) were handled appropriately.
- Missing values in **company** and **agent** were replaced with '0' to indicate direct bookings.
- Missing values in **country** were replaced with 'UNK' for unknown locations.

3.2 Correcting Data Types

- Modified categorical variables such as `hotel`, `arrival_date_month`, `meal`, `country`, and `market_segment` to appropriate `VARCHAR` formats.
- Converted `reservation_status_date` to `DATE` format.
- Changed `adr` (average daily rate) to `FLOAT` for precise calculations.

3.3 Addressing Inconsistent & Incorrect Data

- Standardized country codes (`CN` changed to `CHN` to maintain consistency with ISO 3166-1 standards).
- Merged `Undefined` meal type with `SC` (self-catering).
- Corrected `Undefined` values in `distribution_channel` by assigning them to `Direct`.
- Removed a negative ADR value, as it was invalid.
- Deleted rows where there were no adults but children or babies were present.
- Created an `arrival_date` column by combining `arrival_date_year`, `arrival_date_month`, and `arrival_date_day_of_month`.
- Ensured that repeated guests had valid booking histories.
- Verified that `is_canceled` bookings had an appropriate reservation status.
- Removed bookings with zero stay duration and zero ADR, as they provided no valuable insight.
- Adjusted cases where `required_car_parking_spaces` exceeded the number of adults to align with expected values.

4. Exploratory Data Analysis & Insights

4.1 Booking Trends

- **Most Booked Hotel:** City hotels received more bookings than resort hotels.
- **Seasonality:** Peak booking months were between July and August, while bookings were lower between November and January.
- **Lead Time Analysis:** Most customers arrived at the hotel within 0-30 days of booking, but there were cases with significantly longer lead times.
- **Customer Type:** Transient customers represented 82.6% of bookings, while group bookings were the lowest.
- **Room Type Distribution:** Room Type A was the most reserved (64.7%) but had a lower assignment rate (53.2%), indicating room reassignments or upgrades.

4.2 Cancellation Analysis

- **98.8% of bookings required no deposit, contributing to a 26.5% cancellation rate.**

- **Online Travel Agents (OTAs) had the highest cancellation rates**, showing reliance on third-party booking platforms.
- **Bookings with longer lead times had higher cancellation rates.**
- Higher ADR categories experience more cancellations, indicating that **expensive bookings are more likely to be canceled.**
- Cancellation rates from very short stays gradually **increase from <20% to about 100% at ~30 nights.**
- **Early booking does not always guarantee a lower rate.**

4.3 Revenue & Pricing Insights

- **Seasonal Revenue Trends:**
 - Both City and Resort Hotels experience an upward trend from January with their highest revenue peaks during **July and August**, indicating strong seasonal demands then dips down the remaining months of the year
- **Resort Impact on Resort vs. City Hotels:**
 - Resort Hotels benefit more from extended stays, as guests on vacation tend to stay longer, increasing overall revenue.
 - City Hotels have more short-term business travelers, meaning high turnover and frequent bookings contribute to revenue instead of long stays.
- **Effect of Stay Duration on Revenue:**
 - Longer stays contribute more to total revenue since guests occupy rooms for extended periods. However, the revenue per night may decrease due to discounts offered for extended stays.
 - Short stays have higher price variability per night but may not maximize overall revenue compared to long bookings even though they show stable and lower ADR.
- **Guests with special requests tended to pay higher ADR values**, indicating potential premium service opportunities.

4.4 Customer Behavior & Market Segments

- **Most bookings (59.3%) were through OTAs**, while direct bookings were only 13.5%.
- **Corporate and group bookings were minimal (5.4%)**, highlighting a potential business expansion opportunity.
- **Repeat guests had lower cancellation rates**, suggesting that loyalty programs could improve retention., but they were made up of by a very low percentage of the total guests
- **Room Changes:** A significant percentage of guests did not get the same room they originally booked.
 - Mismatches in room allocation may affect customer satisfaction and reviews.
-

4.5 Regional Guest Insights

- **Portugal had the highest number of guests**, followed by the UK, Spain, and France.
- **City hotels attracted more guests from France and Germany**, while Resort Hotels had higher guests from the UK.

5. Business Recommendations

5.1 Reducing Cancellations

- Implement stricter cancellation policies for bookings with long lead times.
- Offer incentives (e.g., discounts, loyalty points) for non-cancelable bookings.
- Partner with OTAs to reduce last-minute cancellations.

5.2 Revenue Optimization

- Adjust pricing based on seasonal demand using dynamic pricing models.
- Encourage direct bookings through promotional discounts and membership programs.
- Optimize special request handling to enhance customer satisfaction.

5.3 Customer Retention Strategies

- Implement loyalty programs to encourage repeat bookings.
- Personalize marketing strategies for frequent travelers.
- Enhance the guest experience by improving room assignment accuracy.

5.4 Operational Efficiency

- Improve demand forecasting to manage inventory and staffing levels better.
- Focus marketing campaigns on peak booking periods and target low-season travelers with special offers.
- Invest in data analytics for customer segmentation and predictive modeling for better decision-making.

6. Conclusion

This EDA provides valuable insights into hotel booking behaviors, cancellations, and revenue trends. By implementing the recommended strategies, hotels can enhance customer satisfaction, optimize revenue, and improve overall operational efficiency.