Unlocking Insights: EDA for Hotel Bookings

Executive Summary

This report presents the findings of an exploratory data analysis (EDA) conducted on hotel bookings data. The analysis aims to provide insights into customer behavior, booking trends, and potential areas for operational improvement. The key findings include patterns in lead times, cancellation rates, average daily rates (ADR), and the effects of booking changes.

Table of Contents

1. Introduction

In the hospitality industry, understanding customer booking behavior is essential for optimizing inventory management, pricing strategies, and enhancing customer experience. This report delves into the hotel bookings dataset to uncover actionable insights.

2. Data Cleaning and Preparation

Before diving into the analysis, the data underwent a cleaning process to ensure accuracy and reliability. The following steps were taken:
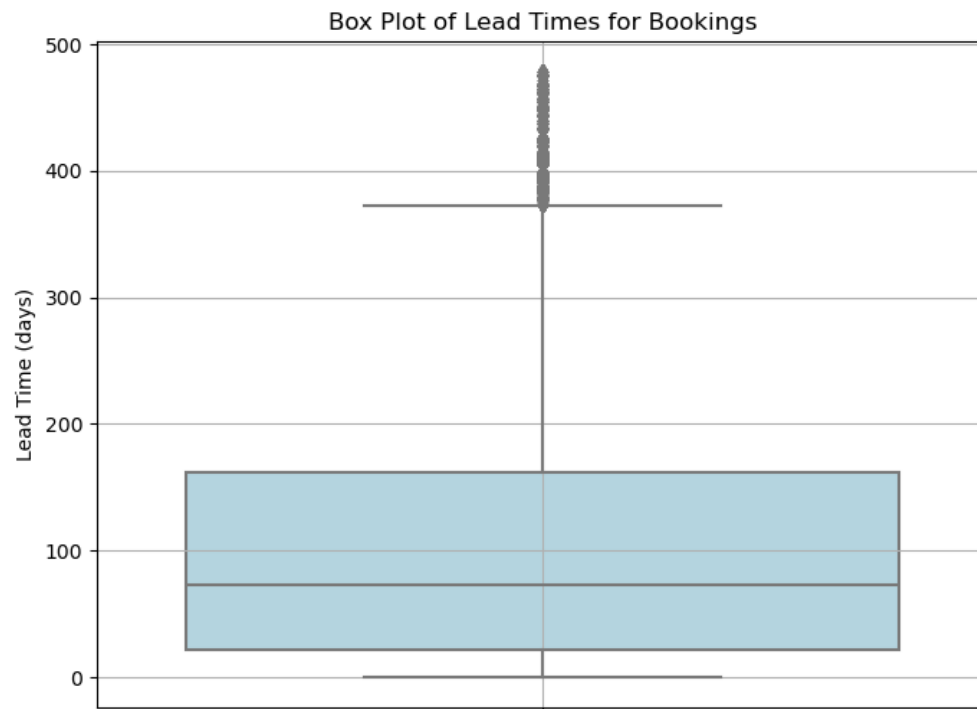
- Dropped columns with a high percentage of missing values (company and agent).
- Removed rows with missing values.
- Handled outliers using Z-scores to maintain data integrity.

3. Lead Time Analysis

- **Distribution of Lead Times for Bookings** (Histogram)

Distribution of Lead Times for Bookings

- **Box Plot of Lead Times for Bookings**
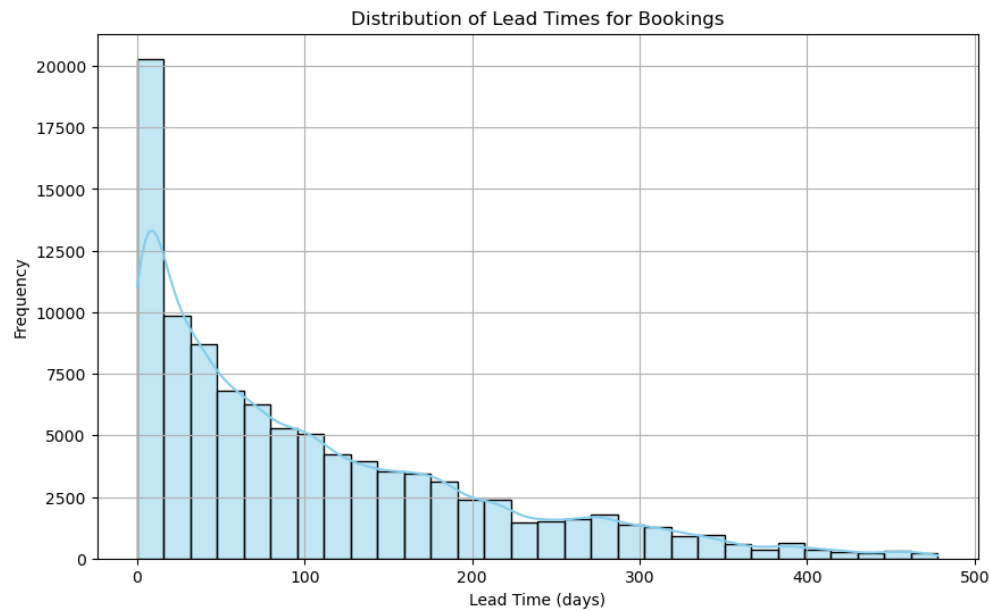


Box Plot of Lead Times for Bookings

Observations and Insights

Based on the analysis of lead time data from hotel bookings, we observe the following insights:

- **Booking Behavior Peaks Near Zero Days:** Many hotel bookings occur with very short lead times, indicating a significant portion of customers make last-minute reservations.
- **Decreasing Frequency with Longer Lead Times:** The frequency of bookings decreases as lead time extends beyond approximately 50 days. Customers tend to book closer to their intended stay date, with fewer bookings made well in advance.
- **Median Lead Time:** The median lead time (approximately 73 days) provides insight into typical booking behavior, serving as a reference point where half of the bookings fall below this duration.
- **Outliers with Long Lead Times:** Some bookings have exceptionally long lead times, exceeding 500 days. These outliers may represent special cases, group reservations, or events requiring extensive planning.

4. Cancellation Rates Analysis

- **Cancellation Rate by Customer Type** (Bar Chart)
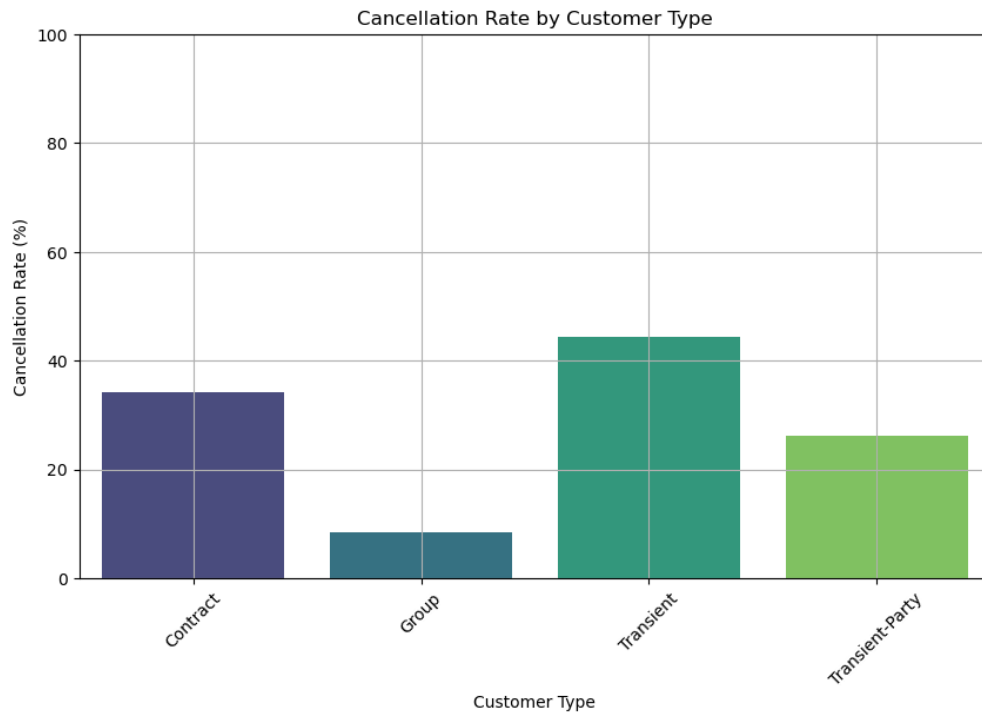
Observations and Insights

Based on the cancellation rates observed:

- **Contract:** Approximately 30% cancellation rate.
- **Group:** Around 10% cancellation rate, significantly lower than others.
- **Transient:** Highest cancellation rate, between 80% and 90%.
- **Transient-Party:** Similar to 'Contract,' around 30% cancellation rate.

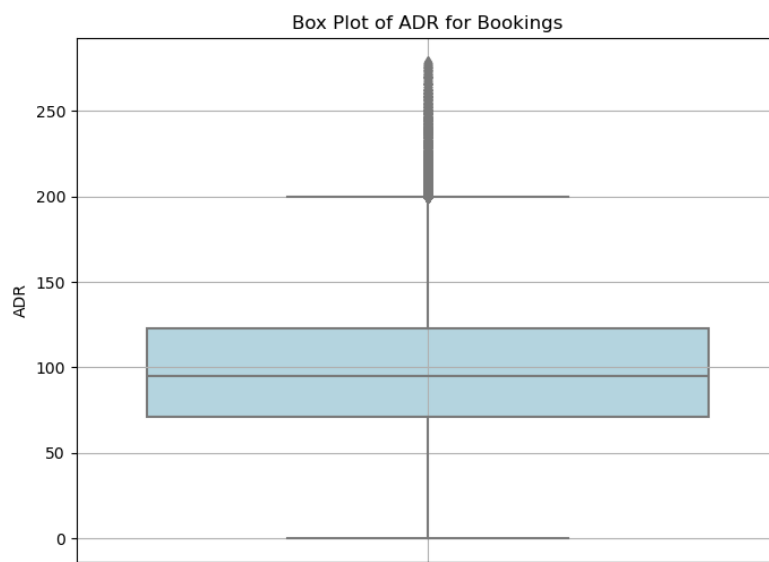From these observations, we gain insights that:

- **Transient customers** are much more likely to cancel their bookings compared to other types.
- **Group customers** show the lowest likelihood of canceling their bookings.

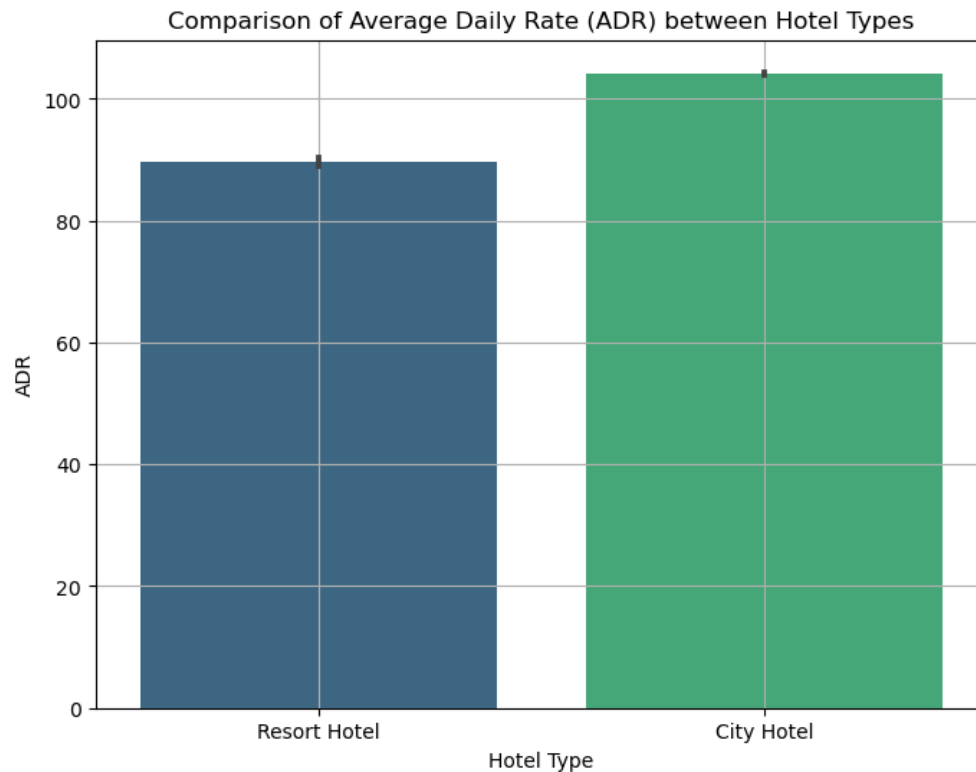# 5. Average Daily Rate (ADR) Analysis

- **Distribution of ADR for Bookings** (Histogram)
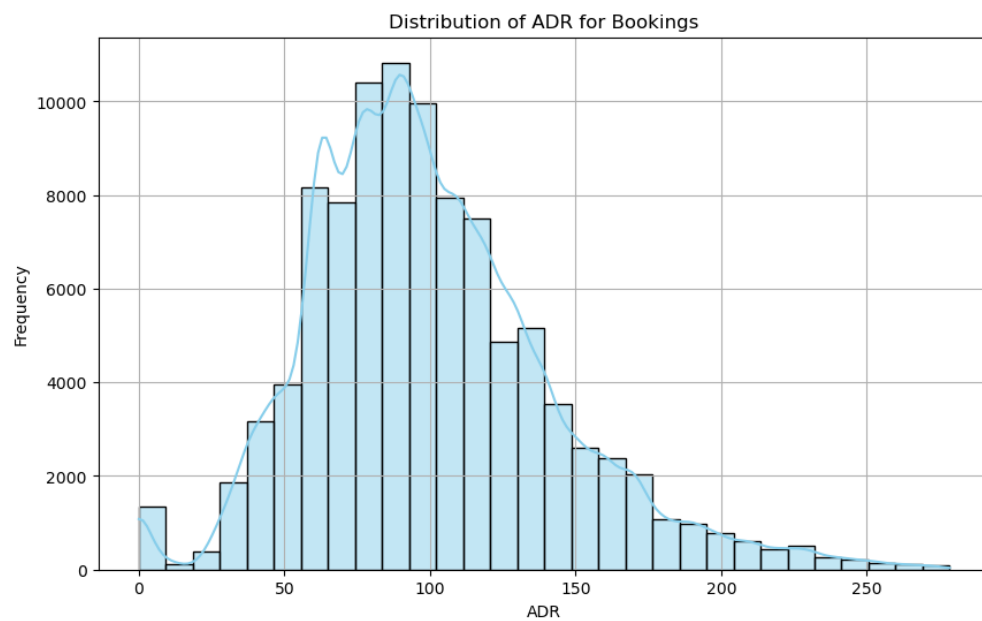
- **Box Plot of ADR for Bookings**

- **Comparison of ADR between Hotel Types** (Bar Chart)

Comparison of Average Daily Rate (ADR) between Hotel Types

Observations and Insights

1. **ADR Distribution:**
   - **Observation:** The majority of bookings have an ADR between 50 and 100 units.
   - **Insight:** Hotels can optimize pricing within this range to attract more customers. High-value outliers might represent premium offerings.
2. **ADR Comparison between Hotel Types:**
   - **Observation:** City Hotels have a higher median ADR (around 100 units) compared to Resort Hotels (around 60 units).
   - **Insight:** City Hotels are generally more expensive, possibly due to location and amenities.

6. Booking Changes Analysis

- **Cancellation Rate by Number of Booking Changes** (Bar Chart)

Cancellation Rate by Lead Time Category
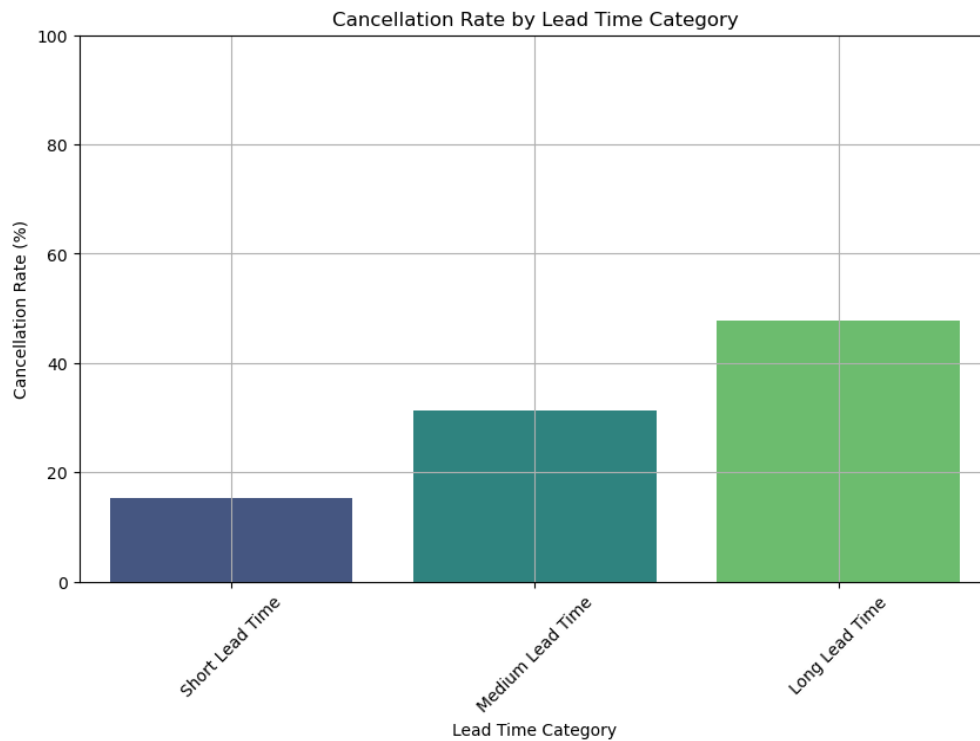
Observations and Insights

Based on the analysis of booking changes:

- **Lower Cancellation with Changes:** Bookings with one or two changes have lower cancellation rates (15.35% and 23.52% respectively) compared to those with no changes (43.29%).

From these observations, we gain insights that:

- **Engagement vs. Cancellation:** The inverse relationship suggests that customers who engage with their booking by making changes are less likely to cancel.
- **Customer Behavior:** Customers who do not make changes might be less committed to their booking, leading to higher cancellation rates.

7. Conclusion

Summary of Findings

The EDA of hotel bookings data has provided several key insights:

- Lead times significantly affect booking and cancellation behaviors.
- Transient customers exhibit the highest cancellation rates, while group customers are more stable.
- City Hotels tend to have higher ADR compared to Resort Hotels.
- Customers making booking changes are generally less likely to cancel their bookings.

Recommendations

Based on these insights, the following recommendations are made:

- **Optimize Pricing:** Focus on pricing strategies within the common ADR range to attract more bookings.
- **Improve Engagement:** Encourage customers to make booking changes as a way to reduce cancellations.
- **Targeted Promotions:** Develop targeted promotions for transient customers to reduce their high cancellation rates.
- **Demand Forecasting:** Utilize lead time patterns to improve demand forecasting and inventory management.