

Analysis Report

**Insights from Hotel Booking Trends:
A Data-Driven Study**

Chapter 1: Introduction

1.1 Background

The hospitality industry has experienced a significant shift in recent years due to digitalization and data availability. Hotels now gather massive volumes of data related to customer behavior, booking patterns, cancellations, seasonal trends, and more. Effectively analyzing this data can help hotels optimize operational efficiency, improve marketing strategies, and enhance customer satisfaction.

This project focuses on performing an **Exploratory Data Analysis (EDA)** on a hotel booking dataset using Python. The analysis aims to uncover patterns and insights that can aid in better decision-making for hotel managers and stakeholders.

1.2 Problem Statement

Hotel bookings can be affected by a wide range of factors such as booking lead time, market segment, customer type, seasonality, and more. Understanding these influencing factors is critical for:

- Reducing cancellations
- Managing overbookings
- Optimizing room allocation and pricing strategies

This analysis seeks to provide a clear understanding of such factors through data-driven visualizations and interpretation.

1.3 Objectives

The main objectives of this project are:

- To clean and preprocess the hotel booking dataset
- To perform univariate, bivariate, and multivariate analysis
- To identify trends and correlations related to booking behavior

- To generate business insights based on visual data exploration
- To provide actionable recommendations for improving hotel operations

1.4 Scope of the Study

This study is limited to the dataset provided, which contains booking information from a city hotel and a resort hotel. The analysis is purely observational and does not involve predictive modeling. However, it serves as a foundation for future work such as demand forecasting or customer segmentation using machine learning techniques.

Chapter 2: Literature Review

2.1 Introduction

The application of data analytics in the hospitality industry is a growing area of research. Hotels are increasingly leveraging data-driven strategies to optimize their operations, understand customer behavior, and enhance revenue management.

2.2 Related Works

Several studies and reports have highlighted the importance of analyzing hotel booking data:

- Research by Antonio et al. (2018) discussed the use of machine learning for predicting hotel cancellations.
- Studies by Guttentag (2015) examined the influence of online reviews and customer preferences in the accommodation industry.
- Business reports from McKinsey and Deloitte emphasize the role of data analytics in improving customer retention and pricing strategies in hospitality.

2.3 Research Gaps

While previous work has primarily focused on predictive modeling or customer segmentation, this project takes a step back to explore the data using EDA techniques. This allows for a more general understanding of trends, seasonality, cancellations, and booking patterns — which can guide deeper machine learning projects in the future.

2.4 Summary

This review reinforces the importance of exploratory data analysis in understanding hotel booking behavior. The insights derived through visual analytics provide a strong foundation for more advanced forecasting and optimization strategies.

Chapter 3: Methodology

3.1 Dataset Description

The dataset used in this project contains detailed information on bookings made at a **city hotel** and a **resort hotel**. It includes variables such as arrival date, lead time, stays in weekend and weekday nights, market segment, distribution channel, repeated guests, booking changes, deposit type, and customer demographics.

- Total records: 119,390
- Features: 32 columns
- Missing values: Present in some features (e.g., **agent**, **company**, **children**)
- Target insight: Cancellation, booking patterns, seasonality

3.2 Tools and Libraries Used

The analysis was performed using the Python programming language with the following key libraries:

- **Pandas** for data manipulation
- **NumPy** for numerical operations
- **Matplotlib** and **Seaborn** for data visualization
- **Jupyter Notebook** for interactive code development and documentation

3.3 Data Cleaning and Preprocessing

The dataset was cleaned and prepared using the following steps:

- Removal of null or missing values from critical columns
- Conversion of date-related columns into proper datetime format
- Dropping irrelevant or duplicate records
- Creating new derived features like total nights, total guests, and month of arrival

3.4 Analytical Techniques Applied

The following types of analysis were performed:

- **Univariate Analysis:** Understanding distributions of individual columns like lead time, hotel type, deposit type, etc.
- **Bivariate Analysis:** Examining relationships between variables such as lead time vs. cancellation, or distribution channel vs. booking count
- **Multivariate Analysis:** Identifying patterns by combining multiple variables (e.g., heatmaps, pair plots)
- **Time Series Trends:** Monthly booking trends, seasonal patterns, and cancellation behaviour over time

3.5 Visualisation Approach

To effectively communicate findings, visualisations were created using:

- Histograms and bar plots for distributions
- Line charts for time series analysis
- Box plots and violin plots for comparing distributions
- Heatmaps for correlation analysis

Each chart is accompanied by an interpretation to support business insight extraction.

Chapter 4: Data Analysis and Visualization

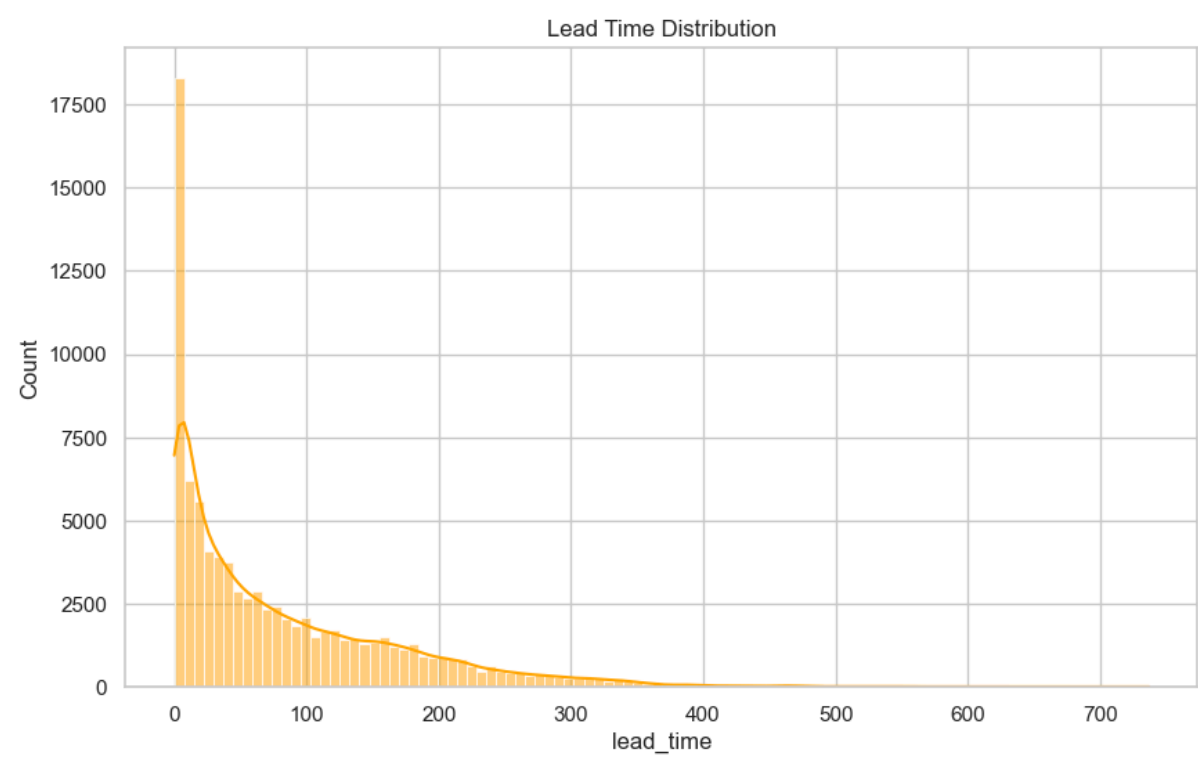
4.1 Overview

In this section, we present the detailed analysis conducted on the hotel booking dataset. The findings are organized into univariate, bivariate, multivariate, and time series analyses. Visualizations accompany each insight to provide a clear understanding of the data.

4.2 Univariate Analysis

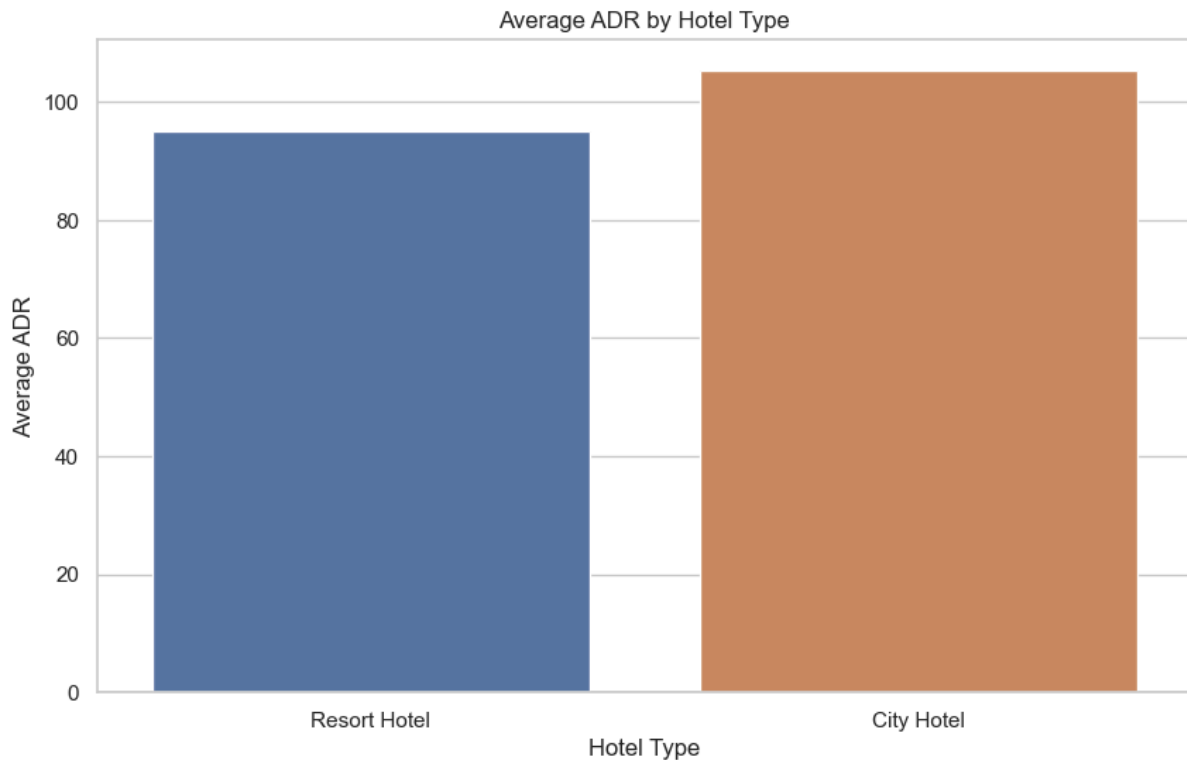
Lead Time Distribution:

A histogram of lead time shows that most bookings are made within 0–100 days before the arrival date, with a peak around 0–50 days. This suggests customers often book close to their stay.



Booking Distribution by Hotel Type:

Bar plots reveal that the city hotel has more bookings than the resort hotel. This may be due to higher business travel in cities.



Deposit Type:

Most bookings have no deposit. Only a small percentage fall into 'Non-Refund' or 'Refundable'.

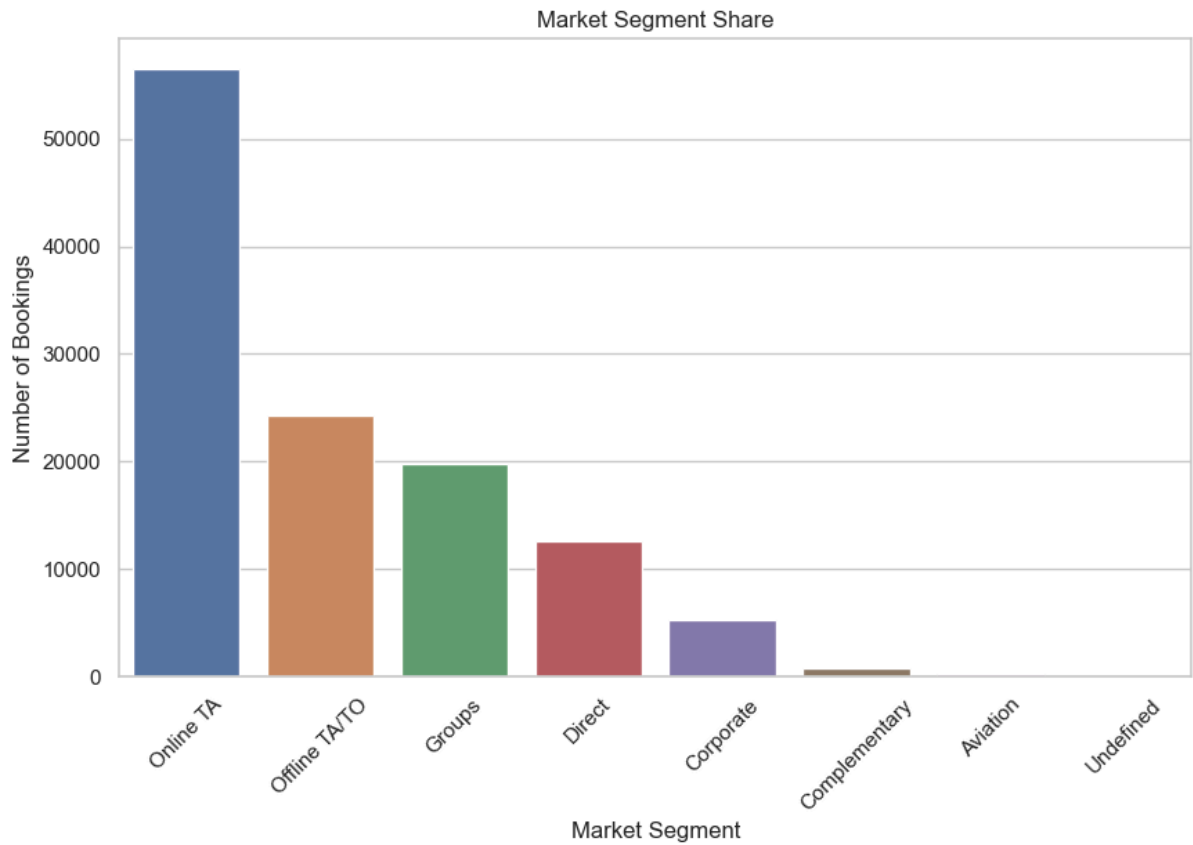
4.3 Bivariate Analysis

Lead Time vs Cancellation:

A boxplot indicates that bookings with longer lead times have a higher cancellation rate.

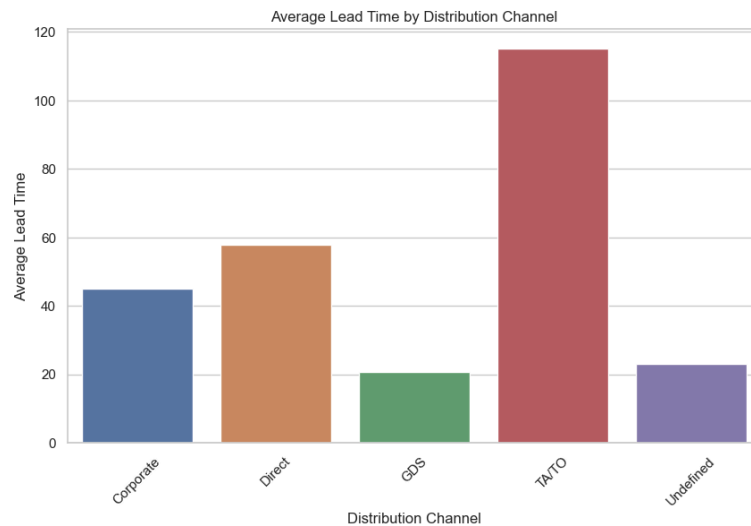
Market Segment vs Booking Count:

The highest number of bookings comes from the 'Online TA' segment, indicating a strong reliance on online travel agents.



Distribution Channel vs ADR (Average Daily Rate):

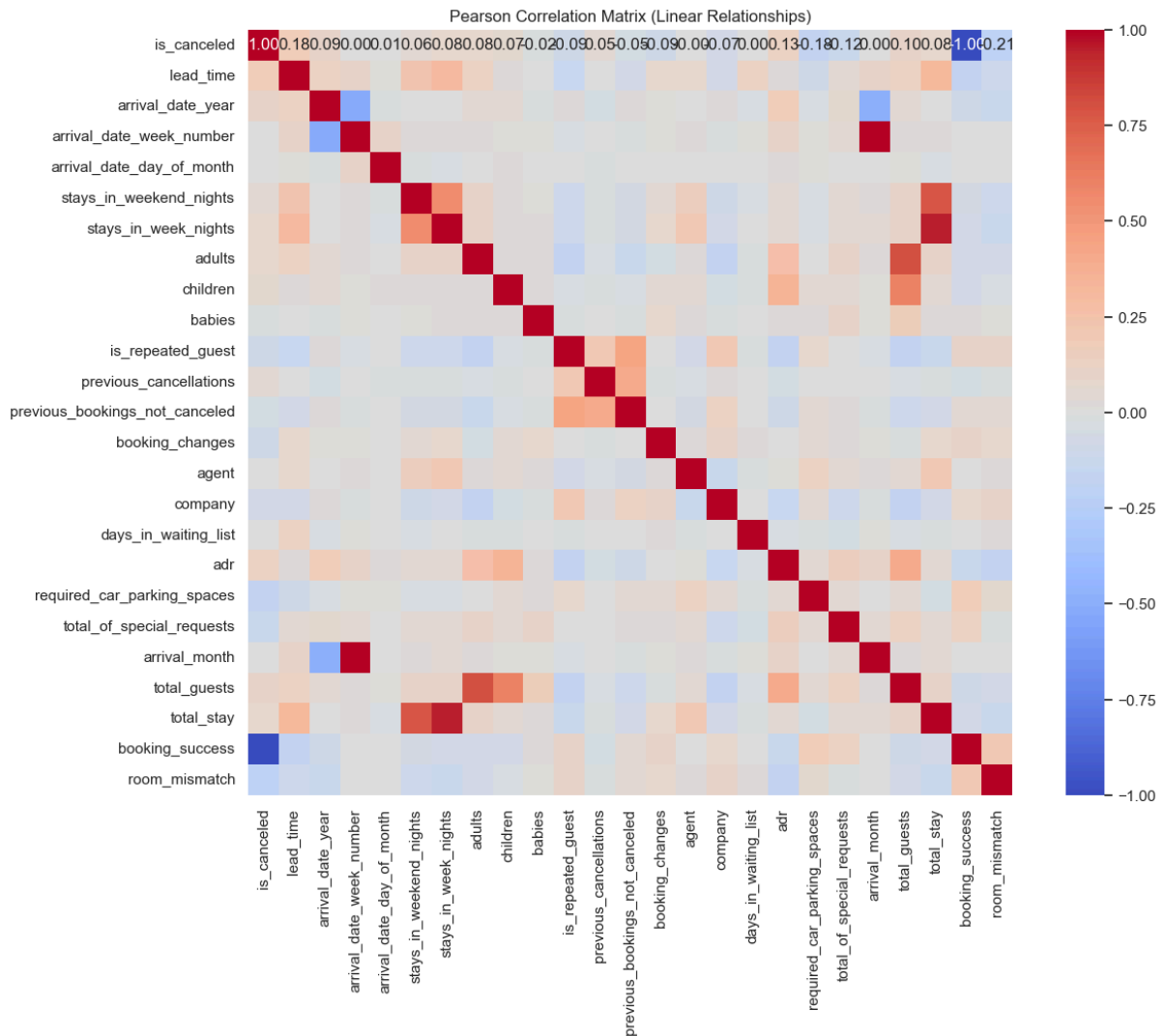
Bookings through corporate channels generally have a higher ADR compared to others.



4.4 Multivariate Analysis

Correlation Heatmap:

A heatmap of numeric variables shows a strong positive correlation between `lead_time` and cancellations, and a negative correlation between `is_repeated_guest` and cancellations.



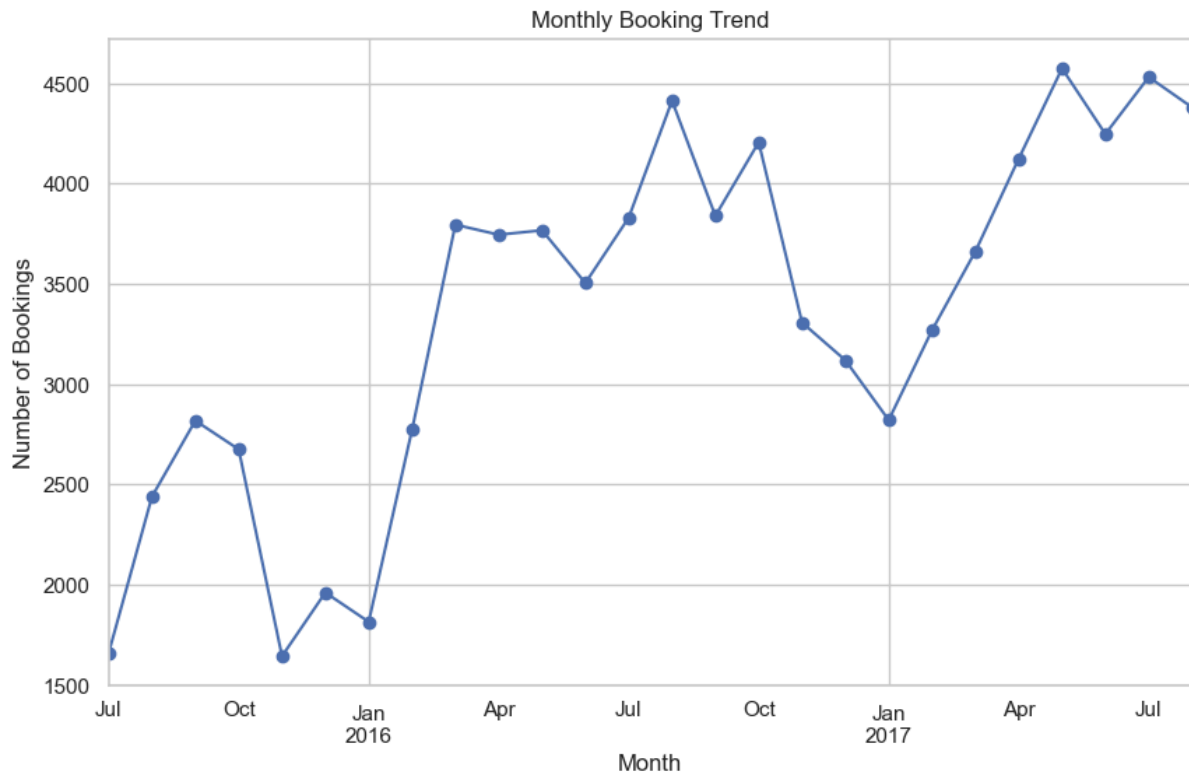
Total Guests vs Total Nights:

Scatter plots indicate that most bookings are for short stays (1–3 nights) with 2 guests.

4.5 Time Series Analysis

Monthly Booking Trends:

Line charts show that bookings peak during summer months (July–August) and drop significantly in winter



Cancellation Trends by Month:

Cancellations also rise during the summer peak season, possibly due to overbooking or last-minute changes.

4.6 Summary of Key Insights

- City hotels attract more bookings than resort hotels.
- Longer lead times are associated with higher cancellation rates.
- Online travel agencies dominate the booking landscape.
- Peak seasons are mid-year (summer), with booking and cancellation spikes.

These insights can guide strategic decisions in marketing, pricing, and customer engagement.

Chapter 5: Business Insights and Recommendations

5.1 Business Insights

Based on the data analysis and visualizations, several key business insights were uncovered:

- 1. High Booking Volume in City Hotels:**
City hotels experience more bookings than resort hotels, likely due to business travel. This suggests the need for different pricing and marketing strategies for each type.
- 2. Lead Time Impacts Cancellation Rate:**
Bookings made far in advance are more likely to be cancelled. Offering incentives or reminders closer to arrival might reduce cancellations.
- 3. Dominance of Online Travel Agents (OTA):**
OTAs are the primary booking channel. Hotels should negotiate better terms with OTAs or promote direct bookings to save commission.
- 4. Seasonal Trends Matter:**
There is a significant spike in bookings and cancellations during summer months. Hotels can manage inventory better by predicting demand.
- 5. Corporate Clients Bring Higher Revenue:**
Corporate distribution channels are associated with higher average daily rates, indicating the importance of maintaining corporate relationships.

5.2 Strategic Recommendations

- 1. Implement a Dynamic Pricing Strategy:**
Adjust pricing based on lead time, season, and booking source to optimize occupancy and revenue.
- 2. Promote Direct Bookings:**
Offer loyalty programs or discounts for direct website bookings to reduce dependency on OTAs.
- 3. Use Predictive Alerts for Cancellations:**
Leverage machine learning in the future to flag high-risk cancellations and prepare accordingly.
- 4. Personalized Marketing:**
Target returning customers or repeated guests with customized offers based on past

Chapter 6: Conclusion and Future Work

6.1 Conclusion

This project successfully applied exploratory data analysis (EDA) techniques to understand the dynamics of hotel bookings using a real-world dataset. Through data cleaning, visualization, and interpretation, several meaningful insights were extracted regarding customer behavior, booking patterns, seasonal trends, and cancellation factors.

Key takeaways include:

- The dominance of city hotels over resort hotels in booking volume.
- A clear link between longer lead times and higher cancellation rates.
- The significant role of online travel agents in the booking ecosystem.
- Seasonal patterns peaking during the summer months.
- The value of corporate channels in generating higher revenue.

These findings provide a data-driven foundation for enhancing hotel operations, marketing strategies, and revenue management.

6.2 Future Work

While this analysis uncovered valuable insights, there are several opportunities for further exploration and enhancement:

1. **Predictive Modeling:**

Future work can include building machine learning models to predict cancellations, estimate booking demand, or segment customers based on behavior.

2. **Customer Segmentation:**

Applying clustering techniques like K-Means can help identify distinct customer groups for targeted marketing.

3. **Text Analytics:**

Incorporating reviews or customer feedback data can provide sentiment analysis and deeper understanding of guest experiences.

4. **Integration with External Data:**

Weather, public holidays, and event data can be combined with booking trends for improved forecasting.

5. **Interactive Dashboards:**

Developing dynamic dashboards using tools like Power BI or Tableau can provide real-time business monitoring and decision support.