Hotel Booking Analysis Report

1. Introduction

This report analyzes hotel booking data to understand patterns, clean inconsistencies, and visualize key trends. The dataset contains information about hotel reservations, including booking details, customer demographics, and reservation status.

2. Data Exploration

2.1 Dataset Overview

- Shape: 119,390 rows × 32 columns

- Data Types:

  - Numerical: 20 columns (int64: 16, float64: 4)

  - Categorical: 12 columns (object type)

2.2 Missing Values Analysis

Several columns contain missing values:

- `children`: 4 missing values

- `country`: 488 missing values

- `agent`: 16,340 missing values

- `company`: 112,593 missing values (removed due to excessive missing data)

2.3 Data Distribution Analysis

Meal Preferences

BB: 92,310

HB: 14,463

SC: 10,650

Undefined: 1,169

FB: 798

Observation: Bed & Breakfast (BB) is the most popular meal plan.

Market Segments

Online TA: 56,477

Offline TA/TO: 24,219

Groups: 19,811

Direct: 12,606

Corporate: 5,295

Observation: Online Travel Agents dominate the booking channels.

Customer Types

Transient: 89,613

Transient-Party: 25,124

Contract: 4,076

Group: 577

Observation: Individual transient bookings are most common.

3. Data Quality Issues & Cleaning

3.1 Identified Issues

1. Missing Values: In children, country, and agent columns

2. Duplicates: 31,994 duplicate rows found

3. Outliers: Detected in `lead_time` and `adr` (average daily rate)

4. Data Inconsistencies:

  - Reservation status not aligning with cancellation status (1,207 cases)

  - Market segment and distribution channel mismatches (20,755 cases)

  - Zero ADR values for non-complementary bookings (1,279 cases)

3.2 Data Cleaning Process

3.2.1 Handling Missing Values

- Removed `company` column (too many missing values)

- Dropped rows with missing `children` and `country` values

- Filled missing `agent` values with code 555

3.2.2 Removing Invalid Data

- Removed bookings with 0 adults

- Eliminated duplicate records

- Removed inconsistent rows where:

  - Reservation status didn't match cancellation status

  - Market segment and distribution channel were mismatched

  - ADR was zero for non-complementary bookings

3.2.3 Outlier Treatment

Lead Time Outliers:

- Used IQR method to identify and remove extreme values

- Before and after comparison shows cleaner distribution

ADR Outliers:

- Applied IQR method to remove extreme average daily rates

- Resulted in more realistic pricing distribution
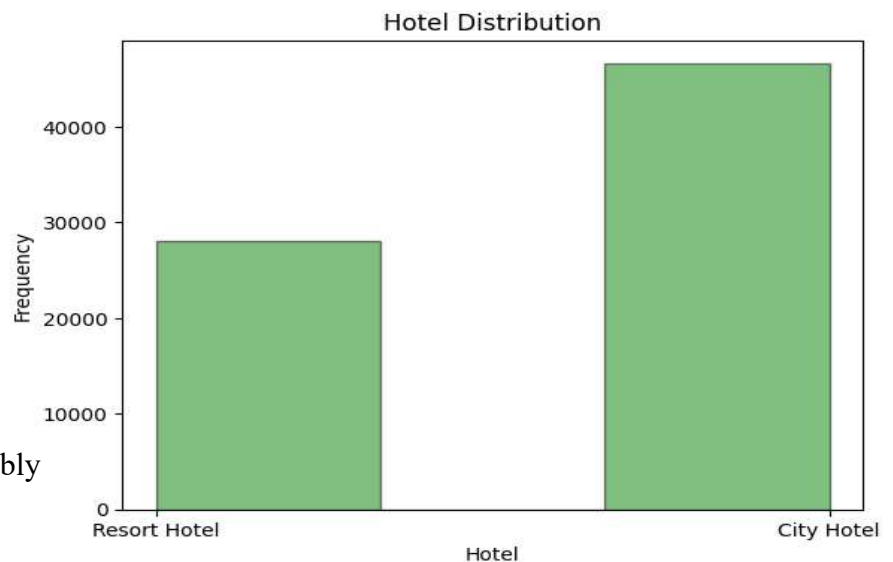
4. Visualizations & Analysis

4.1 Hotel Type Distribution

Key Insights:

- City Hotels dominate the dataset (≈62%)

- Resort Hotels account

for approximately 38% of bookings

- City hotels are more frequently booked, possibly

due to business travel or urban tourism

Hotel Distribution

Percentage Calculation:

- City Hotel ≈ 45,000

- Resort Hotel ≈ 28,000

- Total = 45,000 + 28,000 = 73,000

- **City Hotel**: (45,000/73,000) × 100 ≈ 61.6% → **62%**

- **Resort Hotel**: (28,000/73,000) × 100 ≈ 38.4% → **38%**

City Hotels dominate the dataset, representing around **62%** of the total observations, while Resort Hotels represent around **38%**. This suggests that City Hotels are more frequent, possibly due to their popularity for business trips, urban tourism, or general travel purposes.

## 4.2 Lead Time Analysis

Observations:

- Initial data showed significant outliers in booking lead times

- After cleaning, the distribution became more normalized

- Most bookings are made within reasonable timeframes
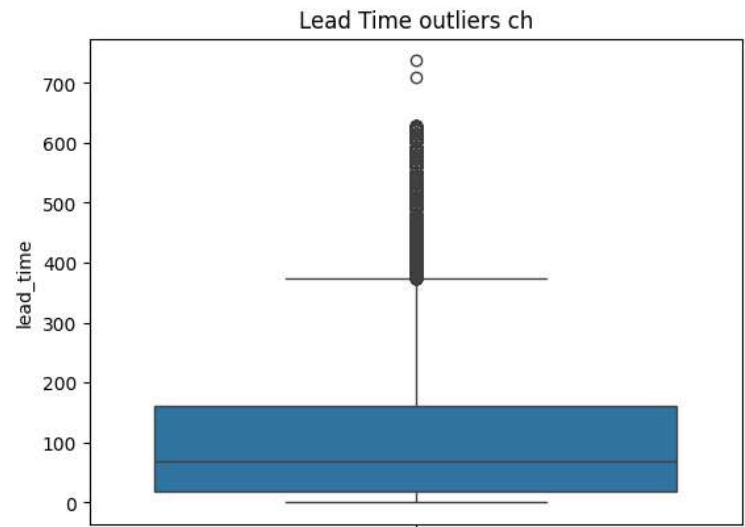
## 4.3 Average Daily Rate (ADR) Analysis

Findings:

- Extreme ADR values were removed through IQR method

- Remaining data shows realistic pricing patterns

- Cleaned data provides better insights into actual booking values
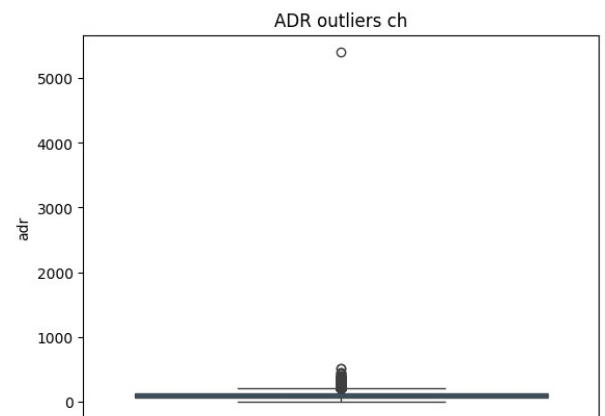
4.4 Overall Cancellation Rate

Analysis: This visualization shows the overall cancellation rate across all hotel bookings, providing a baseline understanding of booking reliability.
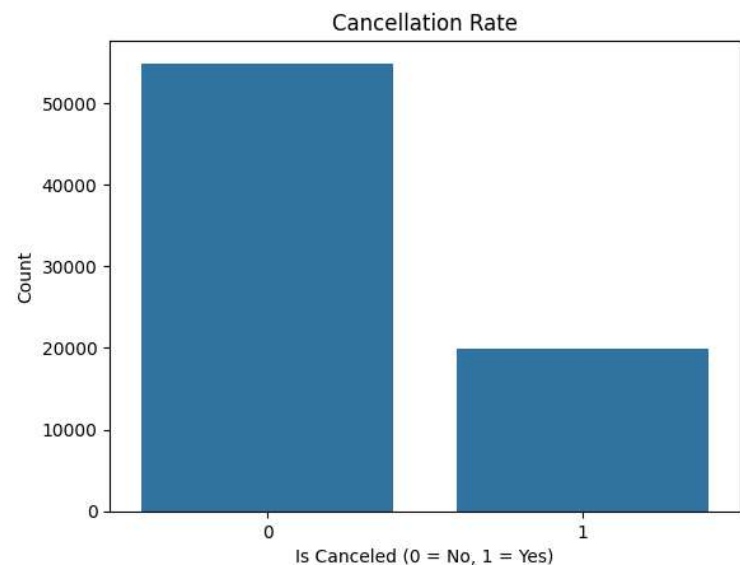
4.5 Reservation Status by Hotel Type

Key Findings:

- **City Hotels** experience more reservations overall but have a higher cancellation rate (29%) compared to Resort Hotels (21%)

Reservation Status by Hotel Type

- This could suggest that City Hotels attract more short-term or business travelers who may be more prone to cancel their plans
- **Resort Hotels** might attract vacationers with more solid travel plans, leading to fewer cancellations
-

4.6 Cancellation Rate by Hotel Type

Quantitative Analysis:

- City Hotels: 29% cancellation rate
- Resort Hotels: 21% cancellation rate
- **Business Implication**: City hotels should implement stricter cancellation policies or require deposits

Cancellation Rate by Hotel Type

4.7 Monthly Reservation Patterns

Seasonal Analysis & Recommendations:

Manage High-Season Demand (July & August):

- **Observation**: August and July have the highest number of reservations, particularly check-outs
- **Action**: Increase staff availability, offer early booking discounts, ensure room maintenance

Reservation Status Per Month

Reduce Cancellations in Off-Season (January & February):

- **Observation**: Fewer total reservations with relatively higher cancellation rates
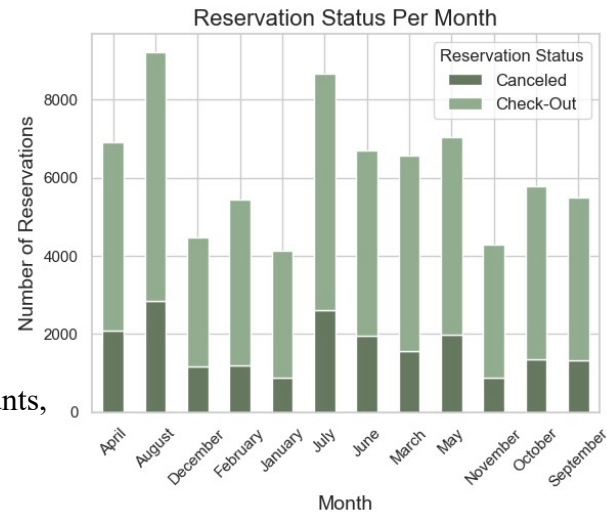
- **Action**: Introduce flexible policies, launch targeted marketing campaigns, partner with travel agencies

Focus on Marketing for Mid-Tier Months (April, June, September):

- **Observation**: Moderate reservation numbers
- **Action**: Promote seasonal packages, create special offers for long stays

**Monitor Cancellation Trends Year-Round**:

- **Observation**: Cancellations occur every month, with notable levels in May and August
- **Action**: Introduce non-refundable rates, offer rescheduling incentives

Improve Guest Retention During December:

- **Observation**: Lower check-outs but moderate cancellations
- **Action**: Create festive packages, collaborate with local attractions

4.8 Customer Type Distribution

Analysis: Shows the breakdown of different customer segments, helping understand the primary customer base.

Distribution of The Customer Type
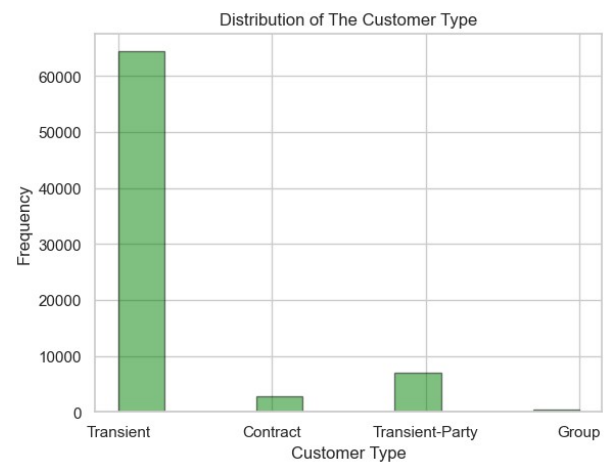
4.9 Family Composition by Room Type

Insights: Reveals which room types are most popular for families with children and babies, informing room configuration strategies.

Number of Adults, Children, and Babies by Assigned Room Type
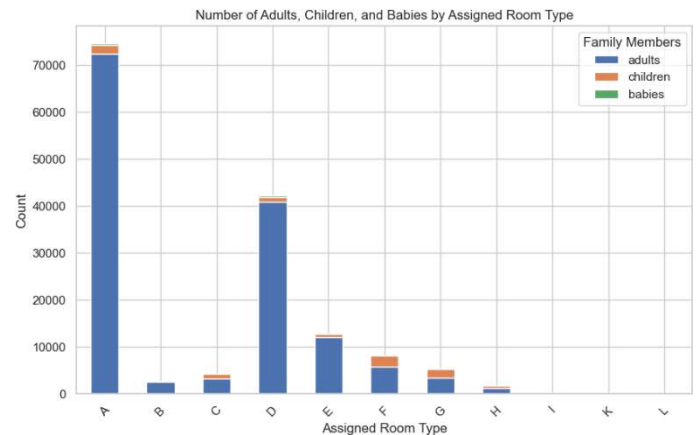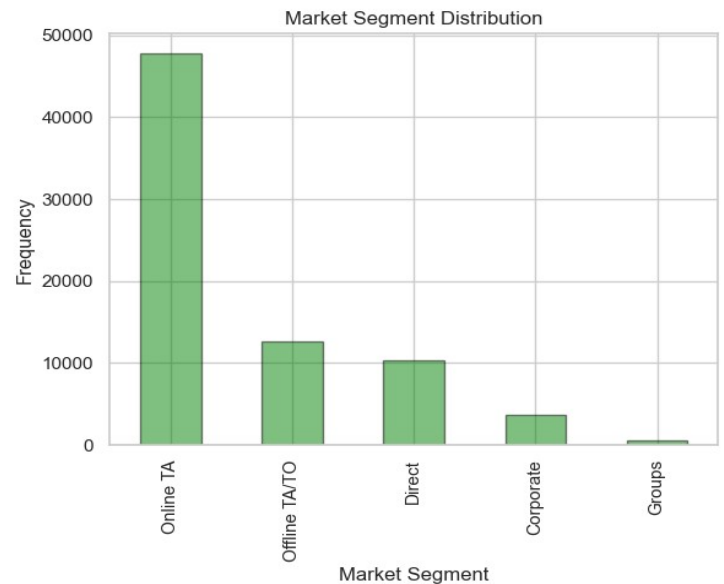
4.10 Market Segment Analysis

Key Findings:

- **Online Travel Agents (OTA)** dominate
the market segment, contributing the highest
percentage of bookings

- This indicates most customers use online platforms
for reservations

Strategic Recommendations:

- **Enhance Online Presence**: Invest in visibility on major OTAs with optimized profiles

- **Exclusive Online Offers**: Provide special deals for online bookings

- **Leverage Customer Reviews**: Encourage positive reviews on OTA platforms

- **Direct Bookings Incentives**: Offer discounts for direct website bookings to reduce commission fees

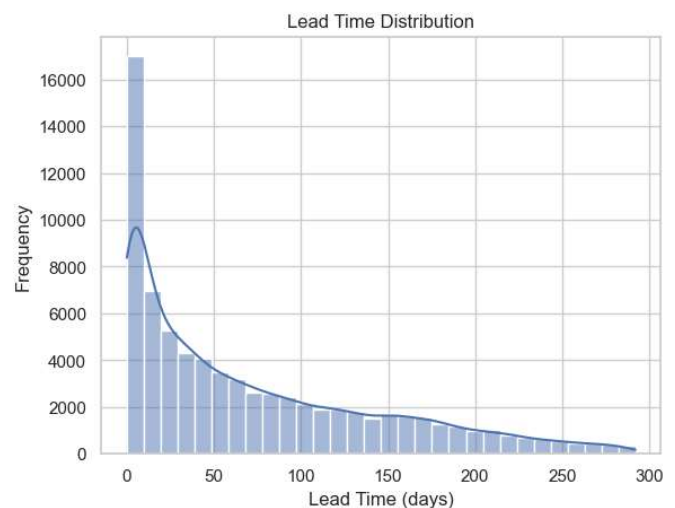## 4.11 Cancellation Rate by Market Segment

Analysis: Identifies which market segments have
the highest cancellation rates, helping target retention
strategies.
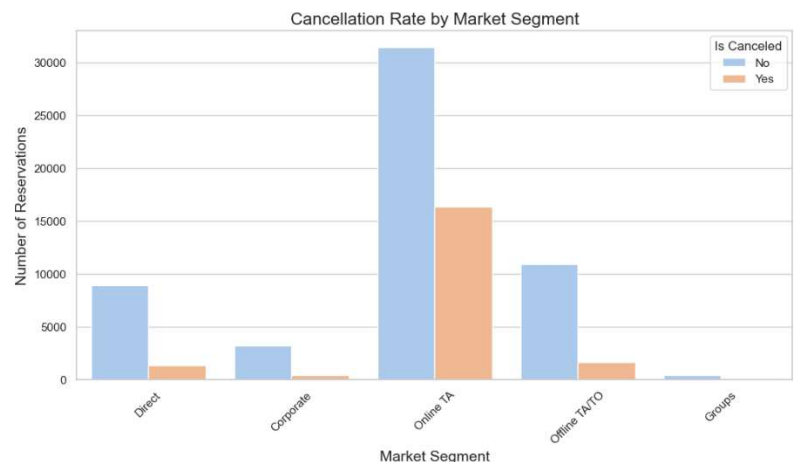
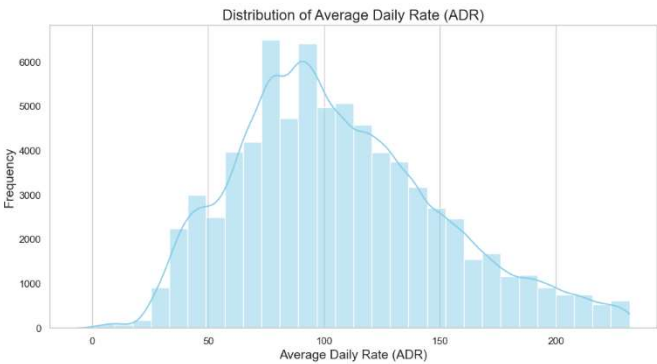## 4.12 Lead Time Distribution

Business Insights:

- Shows how far in advance customers typically book

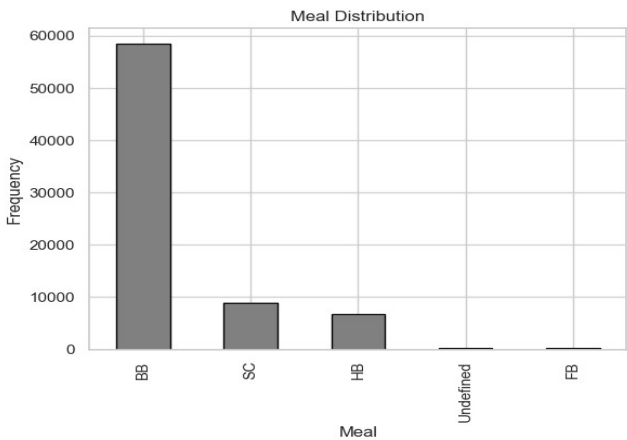- Helps in demand forecasting and resource planning

4.13 Average Daily Rate (ADR) Distribution

Pricing Analysis: Provides understanding of the price point distribution and helps in revenue management strategies.
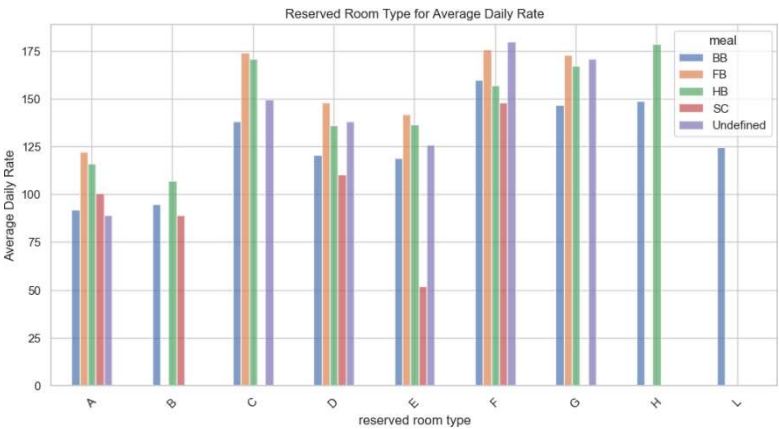

4.14 Meal Plan Preferences

Food & Beverage Insights: Shows customer preferences for meal packages, informing F&B planning and pricing.

4.15 Room Type vs ADR by Meal Plan

Revenue Optimization: Analyzes how different room types and meal combinations affect average daily rates.
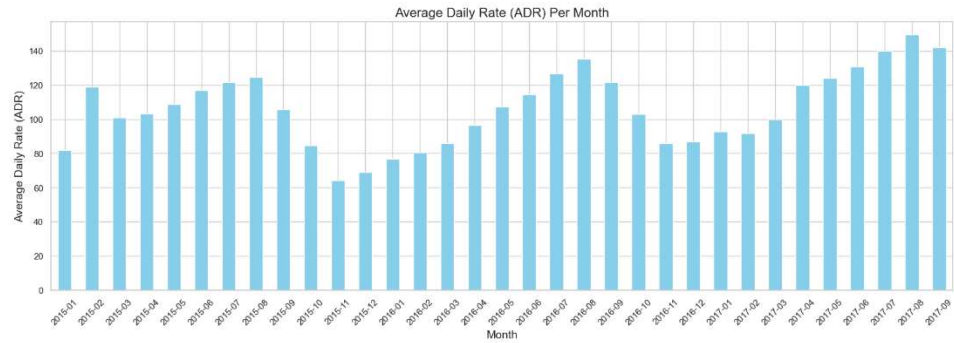

4.16 Special Requests by Hotel Type

Service Quality Insights: Shows which hotel type receives more special requests, indicating customer expectations and service needs.

4.17 Monthly ADR Trends

Revenue Management:

- Shows seasonal pricing patterns
- Helps in dynamic pricing strategies
- Identifies peak revenue periods

Average Daily Rate (ADR) Per Month

5. Data Quality Assessment

5.1 Before Cleaning

- Records: 119,390

- Missing Values: Multiple columns with significant gaps

- Duplicates: 31,994 records

-Inconsistencies: Thousands of logical mismatches

5.2 After Cleaning

- Clean Dataset: Ready for reliable analysis

- Consistent Logic: All bookings follow logical business rules

- Realistic Values: Outliers removed for accurate insights

6. Key Business Insights

6.1 Booking Patterns

- **Channel Dominance**: Online Travel Agents are the primary booking channel (56,477 bookings)
- **Customer Segments**: Transient individual travelers form the largest customer segment (89,613)
- **Hotel Preference**: City hotels receive significantly more bookings than resort hotels (62% vs 38%)

6.2 Cancellation Management

- **Overall Rate**: Significant portion of bookings get canceled
- **Hotel Variation**: City hotels have higher cancellation rates (29%) than resort hotels (21%)
- **Seasonal Patterns**: Certain months show higher cancellation tendencies

6.3 Revenue Opportunities

- **Pricing Strategy**: ADR varies by room type, meal plan, and season
- **Upselling**: Special requests indicate opportunities for premium services
- **Seasonal Pricing**: Clear monthly patterns in ADR suggest dynamic pricing opportunities

6.4 Customer Behavior

- **Planning Horizon**: Lead time distribution shows booking patterns

- **Meal Preferences**: BB (Bed & Breakfast) is overwhelmingly popular

- **Family Travel**: Certain room types are preferred for family stays

7. Strategic Recommendations

7.1 Revenue Optimization

1. **Implement dynamic pricing** based on monthly ADR trends and demand patterns

2. **Create bundled packages** combining popular room types with preferred meal plans

3. **Develop premium offerings** for rooms that command higher ADR

7.2 Cancellation Reduction

1. **Stricter policies** for high-cancellation segments (City hotels, certain market segments)

2. **Non-refundable rate options** at discounted prices

3. **Deposit requirements** for peak season bookings

7.3 Market Segment Strategy

1. **Strengthen OTA partnerships** while developing direct booking incentives

2. **Target underperforming segments** with specialized marketing campaigns

3. **Develop corporate packages** for the contract customer segment

7.4 Operational Planning

1. **Seasonal staffing** aligned with monthly reservation patterns

2. **Inventory management** based on family size and room type preferences

3. **Service enhancement** for high-special-request periods

8. Conclusion

The comprehensive analysis reveals significant insights into hotel booking patterns, customer behavior, and revenue opportunities. The data cleaning process successfully addressed quality issues, ensuring reliable analytical findings. Key takeaways include:

- **Market Position**: Strong presence in city hotel segment with online distribution dominance

- **Revenue Levers**: Clear opportunities in dynamic pricing, package creation, and premium services

- **Risk Management**: Need for better cancellation control, particularly in city hotels and certain market segments

- **Customer Focus**: Understanding of customer preferences across meal plans, room types, and booking channels