

Executive Summary – Hotel Booking Data Analysis

This project presents an in-depth Exploratory Data Analysis (EDA) of hotel booking data to understand reservation patterns, cancellation behavior, pricing dynamics, seasonal trends, and booking channels. The analysis aims to identify the primary factors driving booking cancellations and to provide insights that can help hotels improve revenue management and reduce financial losses.

The cleaned dataset contains **118,897 hotel bookings** across **City Hotels and Resort Hotels**, covering reservation activity from **2015 to 2017**. Data preprocessing included handling missing values, removing columns with excessive null data, standardizing categorical values, converting date fields, and removing extreme ADR outliers to ensure analytical accuracy.

Overall Reservation and Cancellation Overview

- **62.87% of reservations were completed (not canceled).**
 - **37.13% of reservations were canceled**, meaning **nearly 4 out of every 10 bookings did not convert into actual stays.**
 - This high cancellation percentage has a direct negative impact on hotel revenue, occupancy planning, and operational efficiency.
-

Hotel Type Analysis: City Hotel vs Resort Hotel

Booking volume and cancellation behavior differ significantly by hotel type:

- **City Hotels receive the majority of bookings**, indicating higher demand compared to Resort Hotels.
- Cancellation rates by hotel type:
 - **City Hotels:** approximately **42%** cancellation rate
 - **Resort Hotels:** approximately **28%** cancellation rate

- **58% of City Hotel bookings are completed**, while **72% of Resort Hotel bookings are completed**.

Interpretation:

City Hotels experience higher uncertainty, likely driven by business travel, short-term planning, and flexible bookings. Resort Hotels benefit from leisure travelers who plan further ahead and commit to their stays.

Pricing (ADR) and Its Impact on Cancellations

Pricing is one of the strongest drivers of cancellation behavior:

- The **Average Daily Rate (ADR)** is consistently **higher for canceled reservations** than for completed ones.
- Visual trends show that **as ADR increases, cancellation frequency also increases**.
- Monthly ADR analysis confirms that **months with higher room prices experience more cancellations**.

Key Insight:

Customers are highly price-sensitive. Higher room rates increase the likelihood of booking cancellations, especially when alternative options are available.

Seasonal and Monthly Trends

Seasonality plays a major role in booking and cancellation patterns:

- **August** records the **highest number of total bookings and cancellations**, driven by the peak travel season.
- **January shows the highest relative cancellation levels**, despite having lower overall booking volume.
- Both confirmed and canceled bookings peak during high-demand months, but cancellation risk rises significantly during off-peak periods.

Interpretation:

Hotels face different challenges across seasons, including capacity management during peak months and demand stimulation during off-peak months.

Geographic Distribution of Cancellations

- **Portugal (PRT)** accounts for the **largest share of canceled reservations**.
- This aligns with Portugal being the **highest-volume booking country** in the dataset.
- High cancellation volumes in a single country significantly influence overall hotel performance.

Interpretation:

Customer behavior in high-volume markets has a disproportionate impact on cancellation metrics and revenue outcomes.

Market Segment and Booking Channel Analysis

Customer booking behavior varies notably by market segment:

Booking Source Distribution (All Reservations)

- **Online Travel Agencies (OTA): ~47.4%**
- **Offline Travel Agents / Tour Operators: ~20.3%**
- **Group bookings: ~16.7%**
- **Direct bookings: ~10.5%**
- Remaining segments (Corporate, Complementary, Aviation) make up a small share.

Cancellation Distribution (Canceled Reservations Only)

- **Online Travel Agencies: ~46.9%** of cancellations
- **Group bookings: ~27.4%** of cancellations
- **Offline TA/TO: ~18.7%** of cancellations
- **Direct bookings: ~4.3%** of cancellations

Key Insight:

Bookings made through **OTAs and group channels** are significantly more prone to **cancellation**, while **direct bookings** show the lowest cancellation rates.

ADR Comparison: Canceled vs Not Canceled Reservations

Time-series analysis further strengthens earlier findings:

- **Canceled reservations consistently have a higher average ADR** than non-canceled reservations.
- During periods of price spikes, cancellations rise sharply.
- This pattern remains consistent across multiple years.

Conclusion:

Higher prices directly increase cancellation risk, confirming price sensitivity as a dominant behavioral factor.

Business Implications

The analysis clearly shows that hotel booking cancellations are influenced by:

- **Room pricing (ADR)**
- **Hotel type**
- **Seasonality**
- **Booking channels**
- **Geographic concentration of customers**

With an over **37% cancellation rate**, hotels face a serious revenue optimization challenge that can be addressed through targeted, data-driven strategies.

Suggestions and Recommendations

1. Adopt Dynamic Pricing Strategies

Since cancellation rates increase with higher prices, hotels should:

- Use flexible pricing based on demand and seasonality
- Offer early-bird discounts and non-refundable options during high-risk periods

2. Reduce Dependence on OTA Channels

As nearly **47% of cancellations come from OTAs**, hotels should:

- Encourage direct bookings through website discounts or loyalty benefits
- Promote exclusive offers for direct customers

3. Improve City Hotel Retention Strategies

With **~42% cancellation rate**, City Hotels should:

- Introduce stricter cancellation policies during peak demand
- Offer value-added services instead of price reductions

4. Seasonal Campaigns for Low-Demand Months

Given high cancellations in **January**, hotels can:

- Launch targeted marketing campaigns
- Offer bundled packages or extended-stay discounts.

5. Focus on High-Volume Countries

As Portugal contributes the most cancellations:

- Improve service quality and communication in key markets
- Customize offers and policies based on regional customer behavior

Final Conclusion

This Hotel Booking EDA demonstrates that cancellations are **systematic, measurable, and heavily influenced by pricing, booking channels, and seasonality**. The insights derived from visual analysis provide a strong foundation for improving pricing strategy, channel management, and revenue optimization.

By acting on these insights, hotels can reduce cancellation rates, improve occupancy forecasting, and maximize long-term revenue.