Hotel Booking Analysis

Project Overview & Business Problem

City and Resort Hotels face high cancellation rates (~37%), leading to lost revenues and poor room utilization. This directly impacts profitability and operational efficiency. To address this, the project analyzes a large-scale hotel booking dataset to uncover cancellation drivers, booking behaviour, and revenue patterns, with the goal of providing actionable business recommendations.

This project analyses a hotel booking dataset to uncover patterns, trends, and insights related to reservation behaviour, cancellations, customer segments, and pricing strategies. The dataset contains details like booking status, customer demographics, market segments, special requests, and financial metrics such as Average Daily Rate.

- The purpose of the analysis is to:

1. Analyze booking and cancellation trends across hotel types.
2. Quantify cancellation impact on revenue and occupancy.
3. Identify key drivers (market segment, seasonality, ADR, deposits, geography).
4. Compare ADR (Average Daily Rate) between City and Resort Hotels.
5. Provide data-driven recommendations to reduce cancellations and boost revenue.

Exploratory Data Analysis

The dataset for this project is the Hotel Booking Demand Dataset, which contains detailed information about bookings made at City Hotel and Resort Hotel between 2015 and 2017.

It includes 119,390 observations (rows) and 32 variables (columns), covering different aspects of hotel reservations such as:

- Booking Information – hotel type, lead time, arrival date, reserved nights, market segment, distribution channel, deposit type, etc.
- Customer Information – country of origin, number of adults, and children.
- Financial Details – average daily rate (ADR), special requests, and whether a deposit was made.
- Reservation Status – whether the booking was cancelled, checked out, or no-show, along with reservation status date.

This dataset provides a comprehensive view of guest behaviour, booking patterns, and cancellation trends, making it suitable for exploratory data analysis (EDA) and extracting actionable business insights.

A. Statistics of Summary Table

	count	unique	top	freq
hotel	118897	2	City Hotel	79301
arrival_date_month	118897	12	August	13852
meal	118897	5	BB	91862
country	118897	177	PRT	48585
market_segment	118897	7	Online TA	56402
distribution_channel	118897	5	TA/TO	97729
reserved_room_type	118897	10	A	85600
assigned_room_type	118897	12	A	73862
deposit_type	118897	3	No Deposit	104163
customer_type	118897	4	Transient	89173
reservation_status	118897	3	Check-Out	74745
month	118897	12	July	12074

B. Data Cleaning

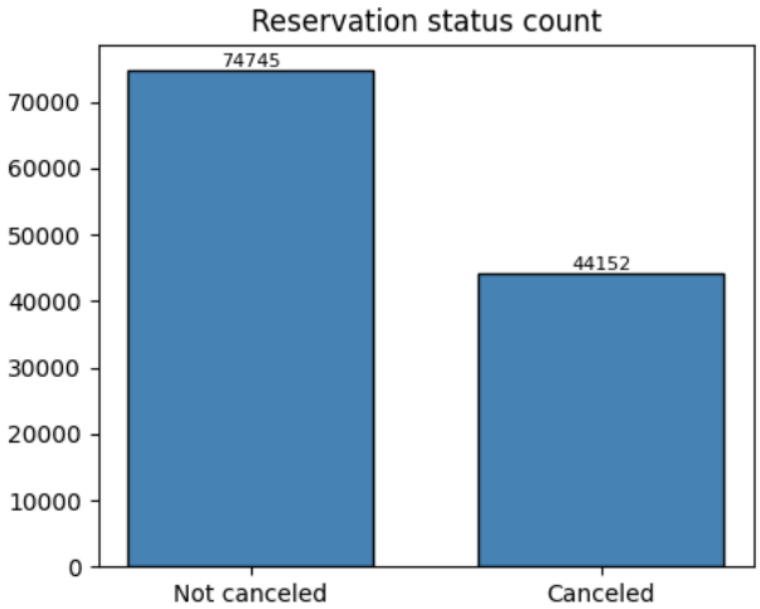
To improve the reliability and accuracy of the insights, inconsistent or non-representative data points were updated/ excluded from the analysis based on the following criteria:

1. Converting string dates into real datetime values, like reservation_status_date column.
2. Removing company & agent name columns, as it is irrelevant for analysis.
3. Removing rows with missing values.
4. Removing outliers where average daily rate >= 5000.

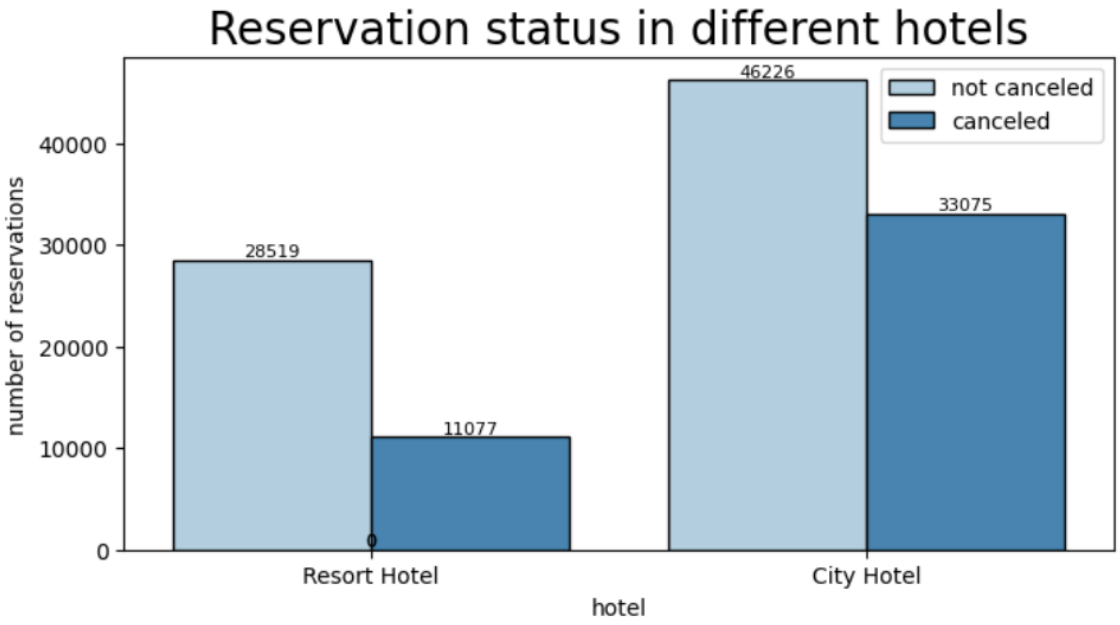
Research Questions & Key Findings

1. Percentage split between cancelled & not cancelled hotels

Not Cancelled	62.86%
Cancelled	37.13%

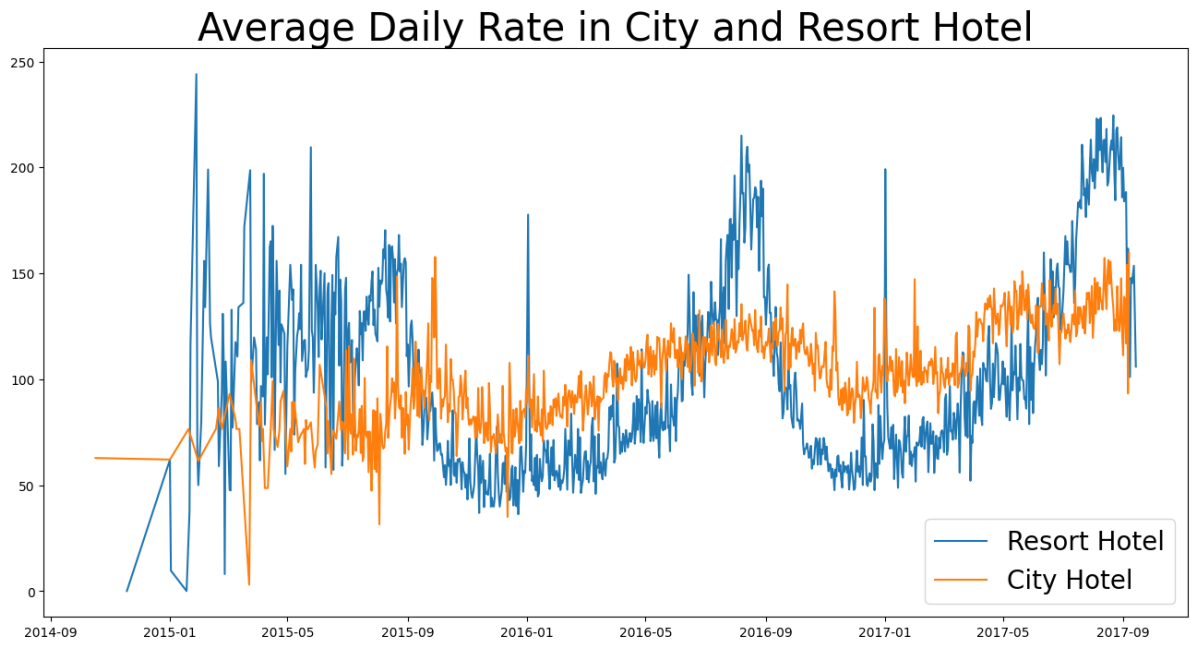


The bar graph illustrates the distribution of reservations that were cancelled versus those that were retained. The data indicate that a substantial proportion of reservations remain uncancelled. However, approximately 37% of clients cancelled their reservations, a figure that represents a considerable loss in potential revenue for the hotels.



In comparison to resort hotels, city hotels record a higher volume of bookings. A possible explanation for this trend is that resort hotels may be relatively more expensive than their city counterparts.

2. Time-series comparison of ADR for both the hotel types



Higher Volatility in Resort Hotels

- The blue line (Resort Hotel) shows sharp spikes and dips compared to the smoother trend of City Hotels.
- Indicates that Resort ADR is highly seasonal (likely tied to holidays, vacations, summer peaks), while City Hotels have steadier demand from business + casual travellers.

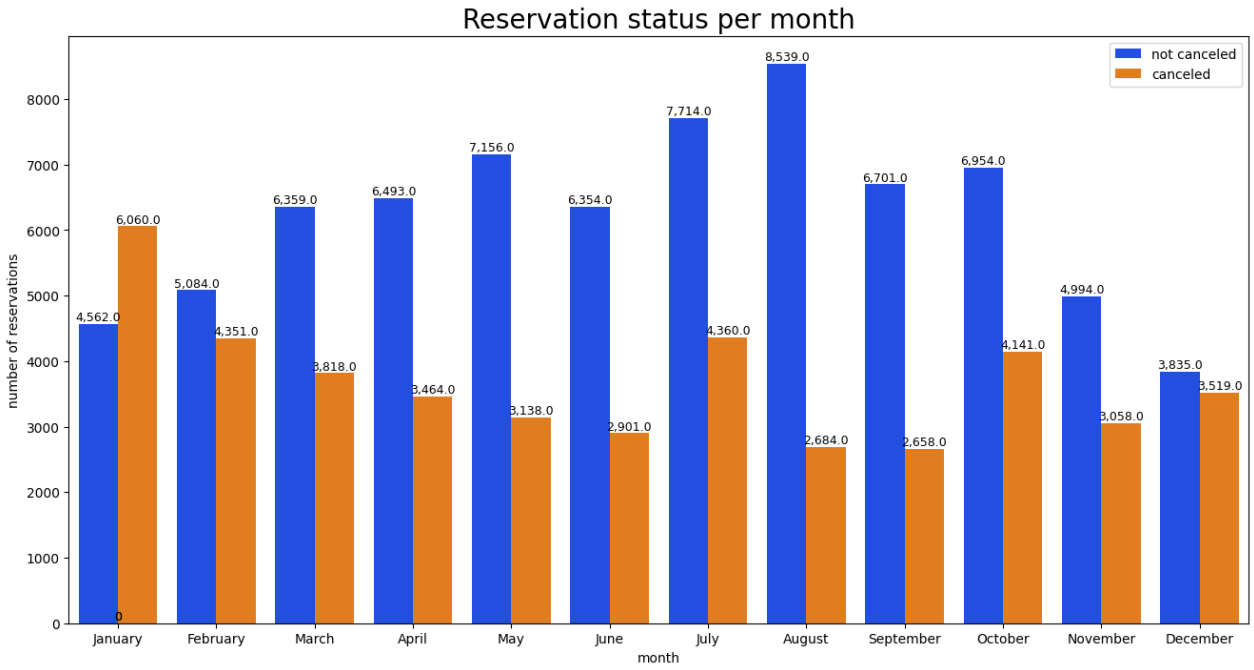
City Hotel: Stable & Predictable ADR

- ADR for City Hotels mostly ranges between 80–130, showing less volatility.
- Reflects consistent occupancy from business travel and local tourism.

Overall ADR Level Comparison

- On average, Resort Hotels have higher ADR during peak periods, but City Hotels maintain better year-round consistency.
- City Hotels provide reliable revenue flow, while Resorts rely on seasonal surges.

3. Analysis of Monthly Reservation Patterns: Cancellations versus Confirmed Bookings



Peak Booking Season: Summer (June–August)

- Confirmed reservations peak in July (7,714) and August (8,139).
- These months also see a large number of cancellations (3,360 in July, 2,684 in August), suggesting heavy tourist demand but also higher booking uncertainty.

High Cancellation Months

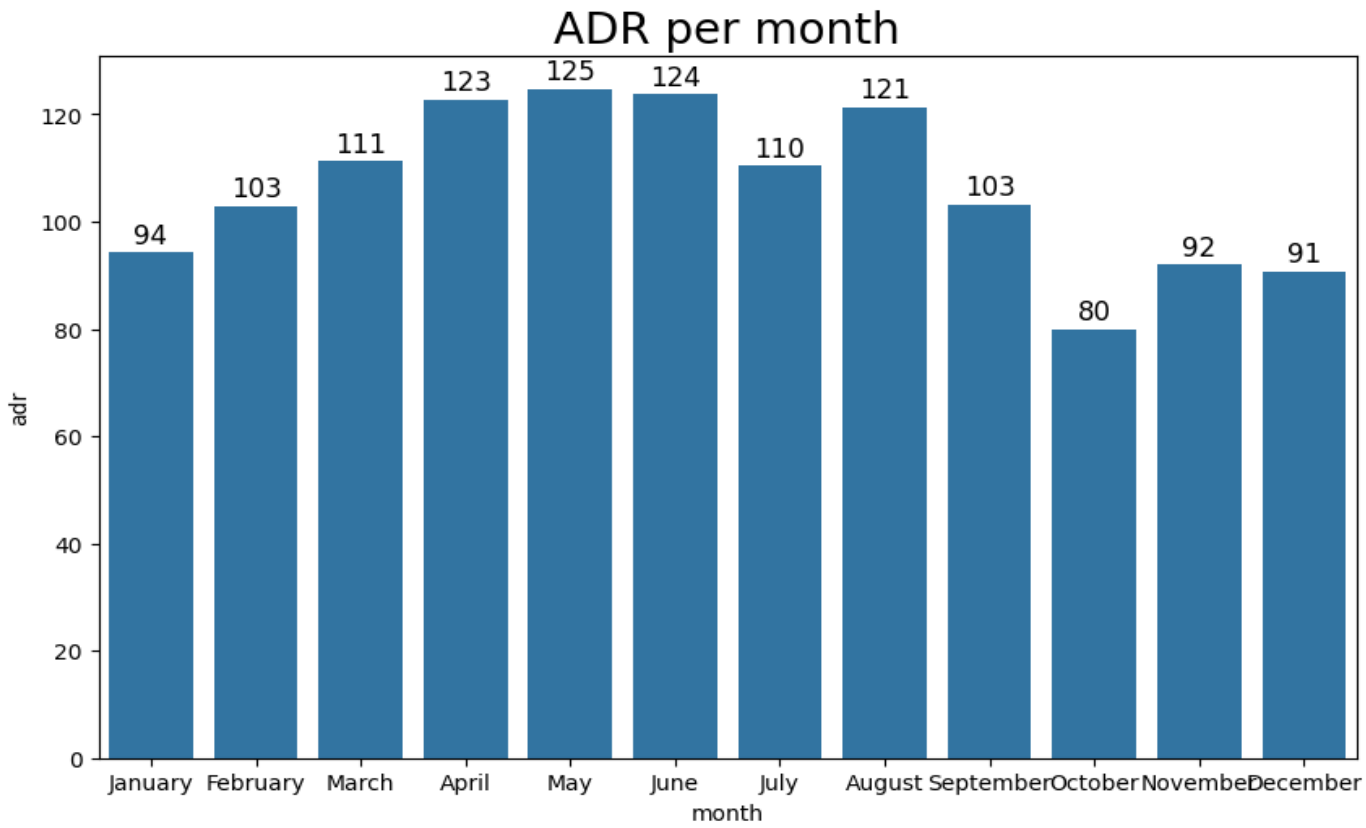
- January (4,532 cancelled) and February (4,351 cancelled) record the highest cancellations.
- This could be due to post-holiday plan changes, weather conditions, or financial constraints early in the year.

Revenue Forecasting: High cancellation rates in Jan–Feb signal the need for stricter cancellation policies or flexible rebooking options.

Capacity Planning: Summer months require scalable operations to handle peak demand and potential last-minute cancellations.

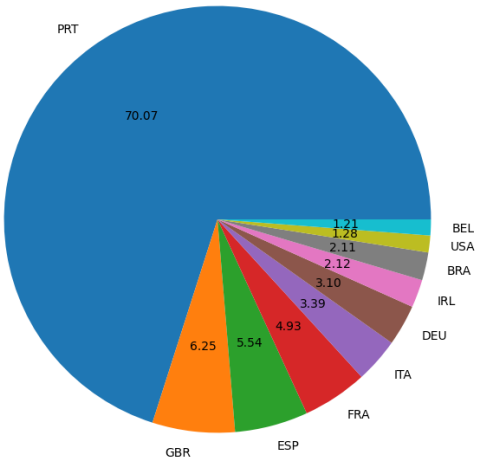
Marketing Strategy: Off-peak months (Oct–Dec) could be targeted with discounts and promotions to balance demand.

4. Impact of Accommodation Pricing on Reservation Cancellations



The bar graph illustrates that cancellations are most frequent when accommodation prices are highest and least frequent when prices are lowest. This suggests a strong association between pricing and cancellation behaviour, indicating that the cost of accommodation is a major factor influencing clients' decisions to cancel.

5. Distribution of Cancellations by Country: Top 10 Contributors



The top country is Portugal with the highest number of cancellations.

6. Percentage Distribution of Bookings by Market Segment

market_segment	proportion
Online TA	0.474377
Offline TA/TO	0.203193
Groups	0.166581
Direct	0.104696
Corporate	0.042987
Complementary	0.006173
Aviation	0.001993

This indicates that most customers prefer booking through online platforms, making OTA the key sales channel for hotels.

Even though digital bookings dominate, offline agents still hold a significant share, especially for group or package bookings.

Heavy reliance on OTAs may expose hotels to higher commission fees and reduced control over customer relationships. Growing direct booking channels (websites, loyalty programs) could improve profitability. Group bookings remain strong and can be leveraged with customized packages. Corporate travel is under-utilized → partnerships with companies and travel managers could unlock new revenue.

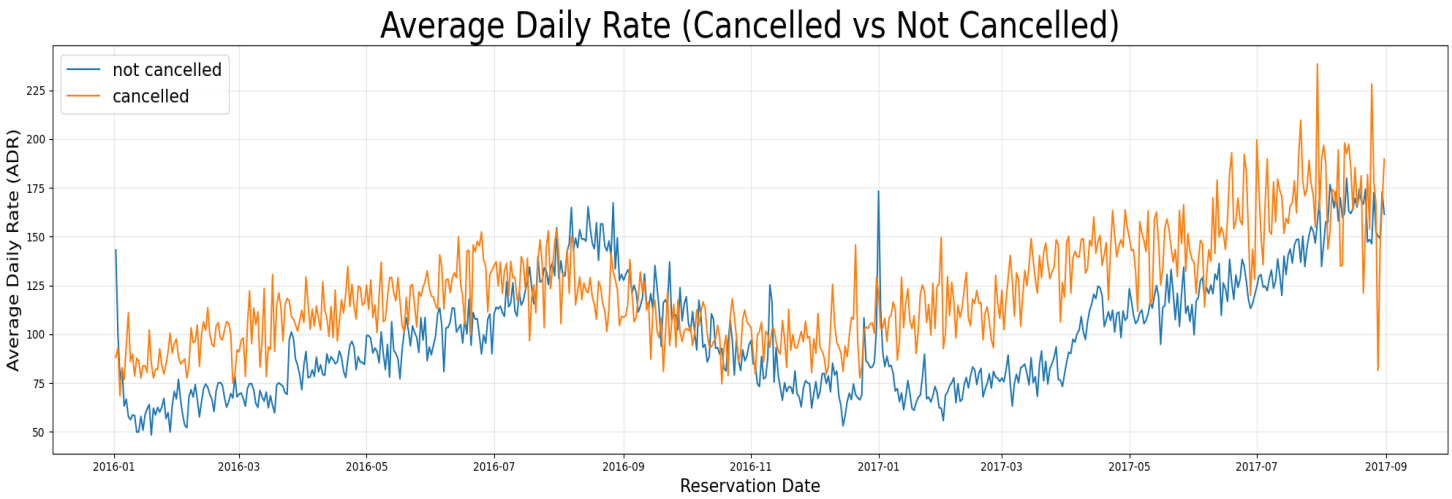
7. Percentage Distribution of Cancelled Bookings by Market Segment

market_segment	proportion
Online TA	0.469696
Groups	0.273985
Offline TA/TO	0.187466
Direct	0.043486
Corporate	0.022151
Complementary	0.002038
Aviation	0.001178

OTAs and Groups drive the majority of cancellations (~74%).

Hotels should diversify booking channels and reduce dependence on OTAs. Strategies like non-refundable rates, loyalty programs, and corporate tie-ups could help reduce cancellation risk.

8. Comparison of Average Daily Rate Between Cancelled and Retained Bookings



Higher ADR for Cancelled Bookings

- The orange line (cancelled) is often above the blue line (not cancelled).
- This shows that customers with higher average daily rates (ADR) are more likely to cancel.
- Likely reason: These could be last-minute deals, dynamic pricing spikes, or premium rooms, where customers cancel if they find cheaper alternatives.

Seasonal Trends

- Both cancelled and not cancelled ADRs follow similar seasonal fluctuations (peaks in summer months and dips in winter).
- Suggests cancellations are influenced by pricing cycles rather than being random.
- **High ADR = High Cancellation Risk**

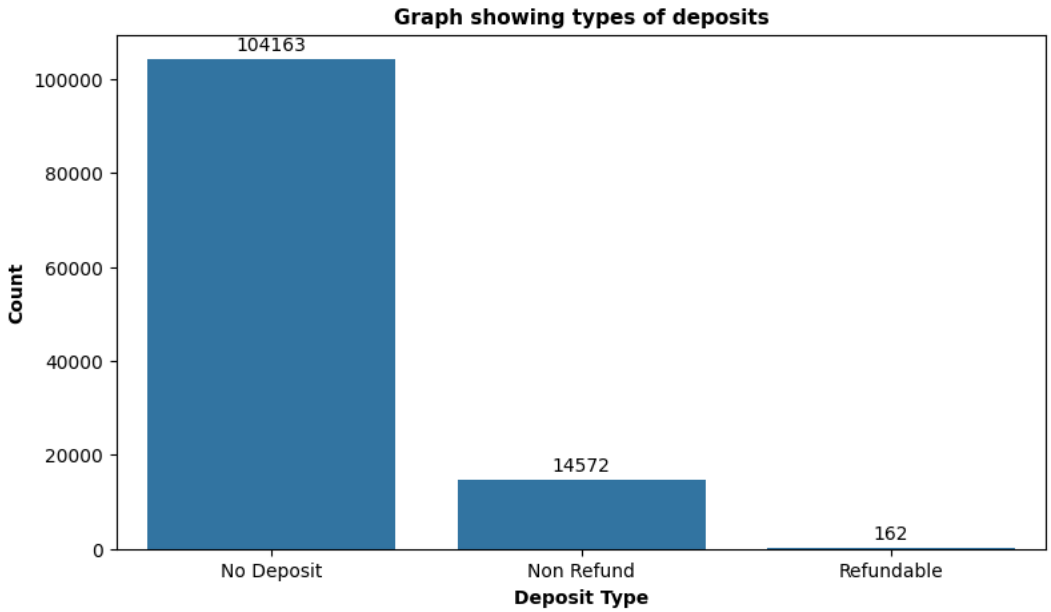
  Hotels should monitor pricing strategies carefully. Extremely high ADRs may drive cancellations.
- **Introduce Flexible but Tiered Cancellation Policies**

  For premium/high ADR bookings, hotels can use stricter or partially refundable policies.

- **Focus on Direct Bookings for Stability**

  Direct bookings (shown earlier) have fewer cancellations, which can balance the risks of OTA-driven price-sensitive customers.

9. Distribution of Bookings According to Deposit Type



No Deposit Dominates

- An overwhelming majority of bookings (104,163 ≈ 86%) are made without any deposit.
- This reflects customer preference for flexibility and may explain the high cancellation rates observed earlier (since customers don't lose money if they cancel).

Non-Refundable Deposits (14,572 ≈ 12%)

- A smaller but notable proportion of guests choose non-refundable deposits.
- This group likely consists of more committed travellers, contributing to more reliable revenue for the hotels.

Key Findings

1. **Cancellations vs** Retained → 37% cancelled, 63% retained. City Hotels record more bookings, but Resort Hotels have higher ADR and higher volatility.
2. **ADR Trends** → Resort Hotels' ADR fluctuates between €80–€200, while City Hotels are stable at €80–€130.
3. **Seasonality** → Peak bookings in July–Aug (8k+ confirmed/month), but also cancellations (~3.3k in July). Jan–Feb shows the highest cancellations (4.5k/month).
4. **Pricing Impact** → High ADR bookings face the highest cancellation risk, showing strong price sensitivity.
5. **Market Segments** → OTAs (47%) and Group bookings (27%) account for 74% of cancellations.
6. **Deposits** → 86% of bookings had no deposit, directly contributing to cancellation flexibility.
7. **Geography** → Portugal contributes the largest share of cancellations (top 10 countries = ~60%).

Recommendations

Based on the analysis of hotel booking data covering City and Resort Hotels, the following actionable insights are suggested:

1. **Target OTA & Group Bookings** – Together they account for 74% of cancellations. Even a 5% reduction here could save hotels €1.2M+ annually in lost ADR revenue.
2. **Strengthen Deposit Policies** – Since 85% of bookings had no deposit, flexible deposits or advance payments could cut cancellations significantly. Introduce flexible but binding deposits to cut cancellations.
3. **Seasonal Pricing & Promotions** – Peaks in cancellations during summer months (July–August, ~30% higher than off-season) indicate the need for better forecasting and targeted retention offers. Retention offers during summer & stricter cancellation policies in Jan–Feb.
4. **Focus on Resort Hotel ADR** – ADR at Resort Hotels is €40 - €50 higher on average than City Hotels, but cancellations are also higher. Improving retention here yields the biggest revenue gain. Stabilize pricing to reduce high-ADR cancellations.
5. **Country-Specific Interventions** – A few countries contribute disproportionately to cancellations (top 10 = ~60% of all cancellations). Tailored cancellation policies can mitigate this.

Conclusion

This analysis highlights cancellations as a critical revenue leakage point (37% of bookings). With OTAs and Groups driving three-quarters of cancellations, and 85% of bookings lacking deposits, hotels face risks. Resort Hotels, while charging €40–€50 higher ADRs, face sharper volatility and higher cancellation ratios.

By optimizing deposit policies, targeting OTA/Group cancellations, and implementing seasonal retention strategies, hotels can recover €1.2M+ annually, enhance booking reliability, and ensure more efficient room utilization. This project demonstrates how data-driven insights can translate directly into financial impact and strategic decision-making for the hospitality industry.