

# Hotel Booking Cancellation Prediction & Reservation Intelligence

A full-cycle hospitality analytics system — City Hotel & Resort Hotel

119,400 reservations · 235,000 guests · 2 hotel types · 12-month analysis

5 Interactive Dashboards

Power BI

ML Cancellation Predictor

Python

Segmentation Queries

SQL

Executive Strategy Report

PDF

 Reduce Cancellations

 Grow Revenue & ADR

 Retain Guests Directly

by MOHD FAIJ

# Data Foundation & Analytical Pipeline

From raw CSV to five dashboards and a machine learning model

119,400

Total Bookings

Full reservation records across both hotel types

79K

City Hotel

Urban property – dominant booking volume

235,000

Total Guests

All arriving guests across the analysis period

40K

Resort Hotel

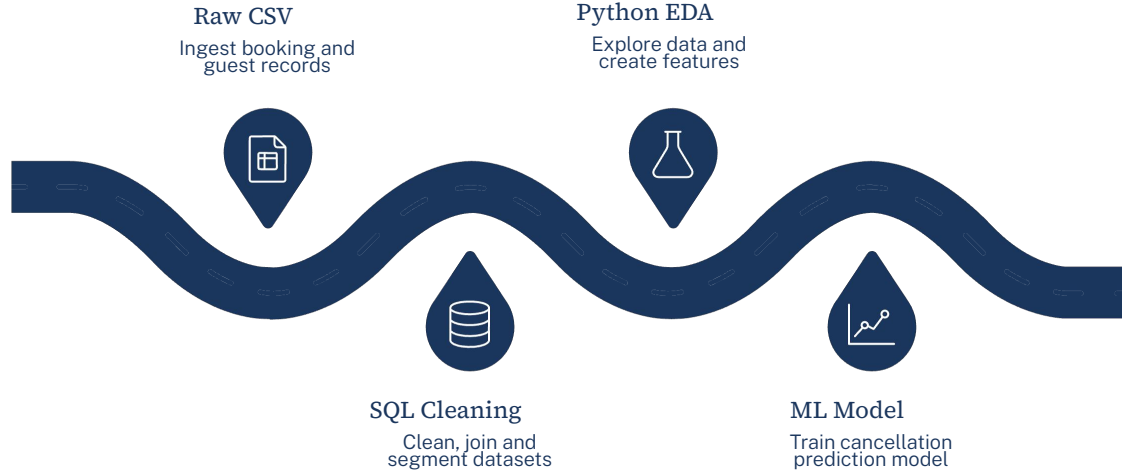
Leisure property – higher ADR & stay duration

## Dataset Features – Left

- Booking channel
- Customer type
- Average Daily Rate (ADR)
- Lead time (days)
- Cancellation status
- Arrival month

## Dataset Features – Right

- Room type (A-E)
- Deposit type
- Special requests count
- Parking requirement
- Previous cancellations
- Booking changes



SQL / PostgreSQL

Power Query M

DAX + Power BI Service

Python + Pandas

Scikit-learn + Jupyter

PERFORMANCE OVERVIEW

## Hotel Performance at a Glance

Bookings, cancellations, ADR, and guest volume across City & Resort Hotel

119,400  
Total Reservations

\$101.83  
Average ADR

3.2%  
Repeat Guest Rate

37.0%

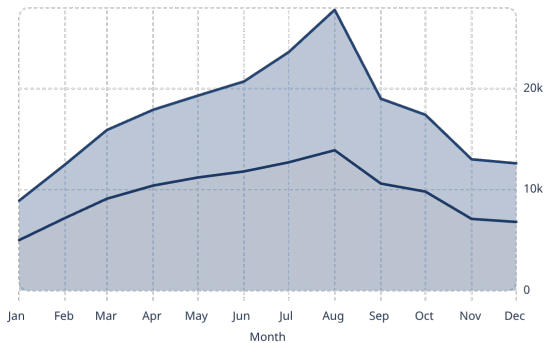
Cancellation Rate  
Critical risk indicator

235,000  
Total Guests

3.4  
Avg. Nights Stay

### Monthly Booking Trend — Total vs Confirmed

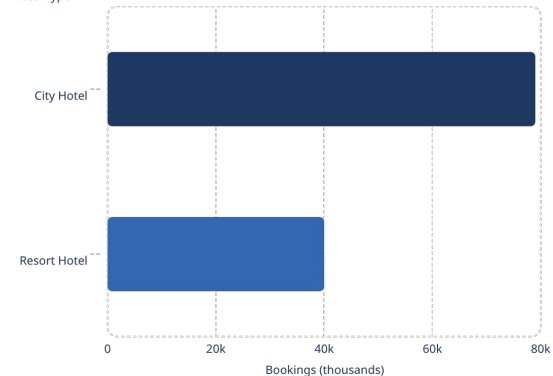
Confirmed Bookings Total Bookings



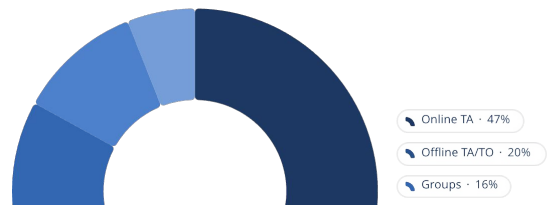
July and August represent the highest revenue opportunity — both booking volume and ADR peak in this window. Peak Season: Jul/Aug at 12.7K-13.9K bookings.

### Reservation Volume by Hotel

Hotel Type



### Booking Share by Market Segment



# Where Revenue Is Being Lost

Segment-level cancellation rates, lead time impact, and loss distribution

44,200

Cancelled Reservations

● Revenue at risk

37.0%

Cancellation Rate

● Above industry norm

75,200

Confirmed Reservations

● Realised revenue base

104

Avg. Lead Time (Days)

14.5%

Returning Guest Cancel Rate

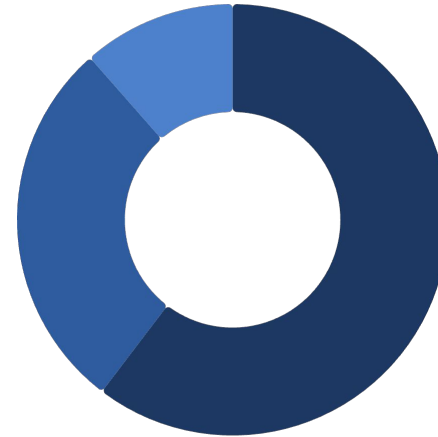
● Significantly lower risk

## Cancellation Rate by Market Segment

Segment	Rate	Risk Level
Groups	61.06%	● High
Online TA	36.72%	● Medium
Offline TA/TO	34.32%	● Medium
Aviation	21.94%	● Medium
Corporate	18.73%	● Low
Direct	15.34%	● Low
Complementary	13.06%	● Low

Groups carry a 61% cancellation rate – the highest single-segment risk. Online TA follows at 36.7%. Combined, these two segments account for the majority of all cancelled revenue.

## Predicted Risk Distribution



● Low Risk (60.3%) · 60.35%

● Medium Risk (28.2%) · 28.16%

● High Risk (11.5%) · 11.48%

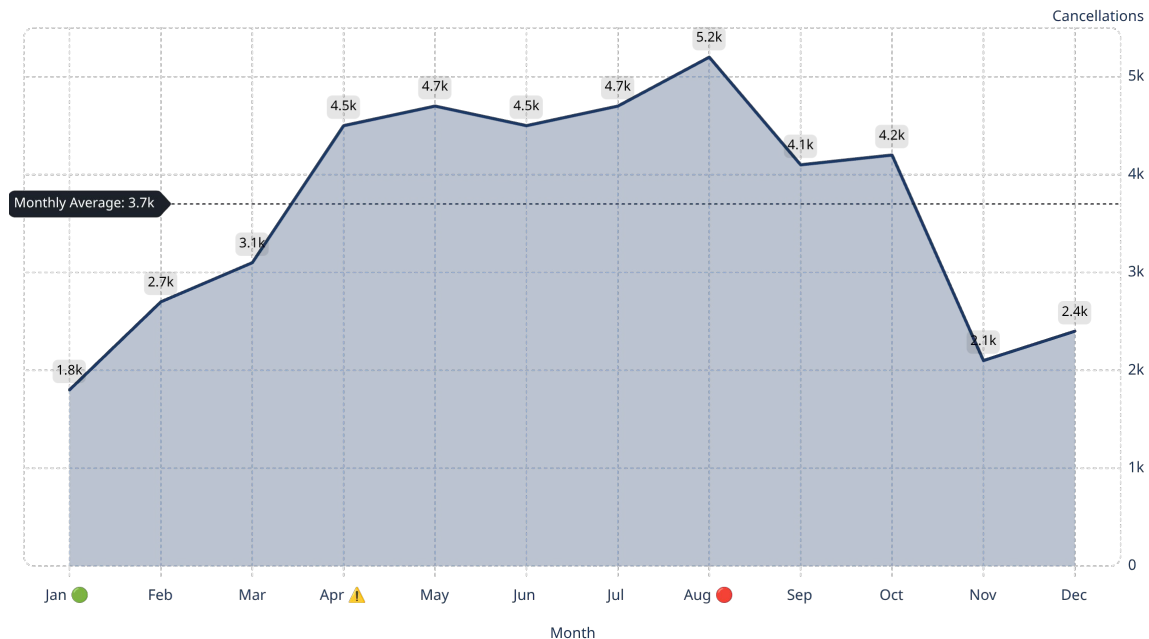
● Low Risk  
72,000 bookings · 60.3%

● Medium Risk  
33,600 bookings · 28.2%

● High Risk  
13,700 bookings · 11.5%

# Cancellations Follow a Predictable Seasonal Pattern

Monthly cancellation volume — January through December. April–October is a sustained 6-month high-risk window averaging 4,500+ cancellations per month.



## August Peak

5,200 cancellations — single highest month, coinciding with peak booking season and highest ADR. Requires aggressive deposit enforcement.

## April–October High-Risk Season

Sustained 6-month window averaging 4,500+ cancellations per month. Revenue protection protocols must be active throughout this window.

## Nov–Dec Natural Drop

Sharp decline to 2,100–2,400 cancellations — lower exposure period, but also lower revenue volume overall.

⊗ Recommendation: Deploy predictive cancellation alerts and mandatory deposit triggers for all bookings during April–October high season.

## Who Pays More – And Who Cancels More

ADR intelligence by hotel type, customer segment, and booking status

\$101.83

Overall ADR

40.7%

Transient Cancel Rate

● Highest risk segment

\$107.0

Transient ADR  
Highest paying – highest risk

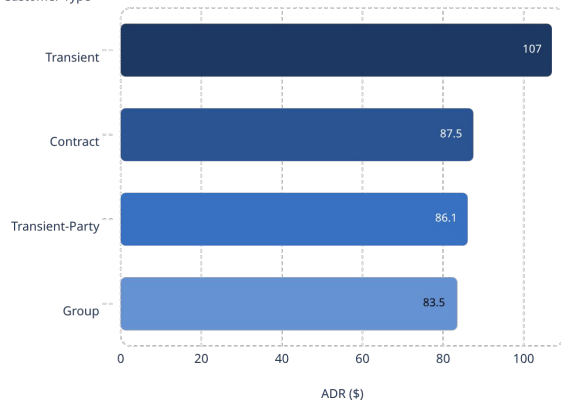
10.2%

Group Cancel Rate

● Lowest cancel rate

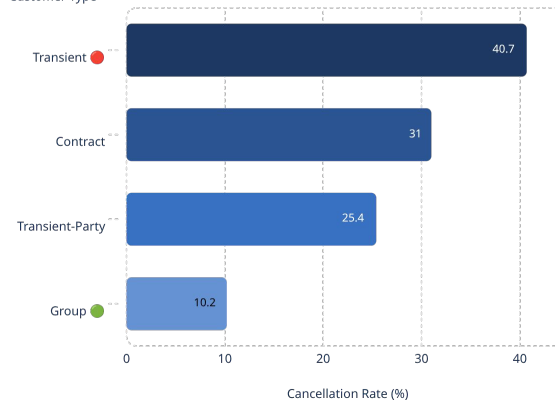
ADR by Customer Type

Customer Type



Cancellation Rate by Customer Type

Customer Type



Key Tension: Transient guests pay the most (\$107) but cancel the most (40.7%) – highest revenue risk segment.

Room Type ADR

Room	ADR	Note
Room C	\$160.2	🔥 Premium – dynamic pricing opportunity
Room E	\$124.5	High yield
Room D	\$120.7	Strong yield
Room A	\$90.8	Highest volume (86K bookings)
Room B	\$90.4	Standard

# Who Books, How They Arrive & How Long They Stay

Stay patterns, booking channels, repeat rates, and room demand by customer type

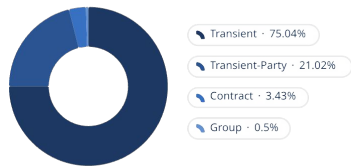
**75.1%**  
 Transient Share  
 89,600 bookings — dominant segment

**3.43**  
 Avg. Stay (Nights)

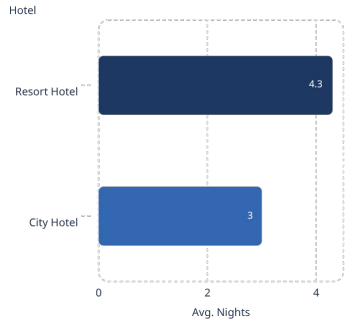
**3.19%**  
 Repeat Guest Rate  
 Loyalty investment opportunity

**12.49%**  
 Room Upgrade Rate

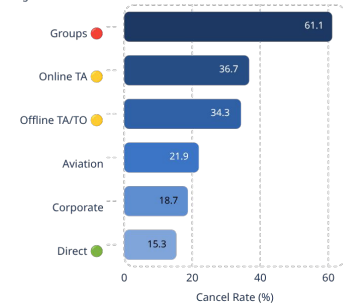
Booking Breakdown by Customer Type



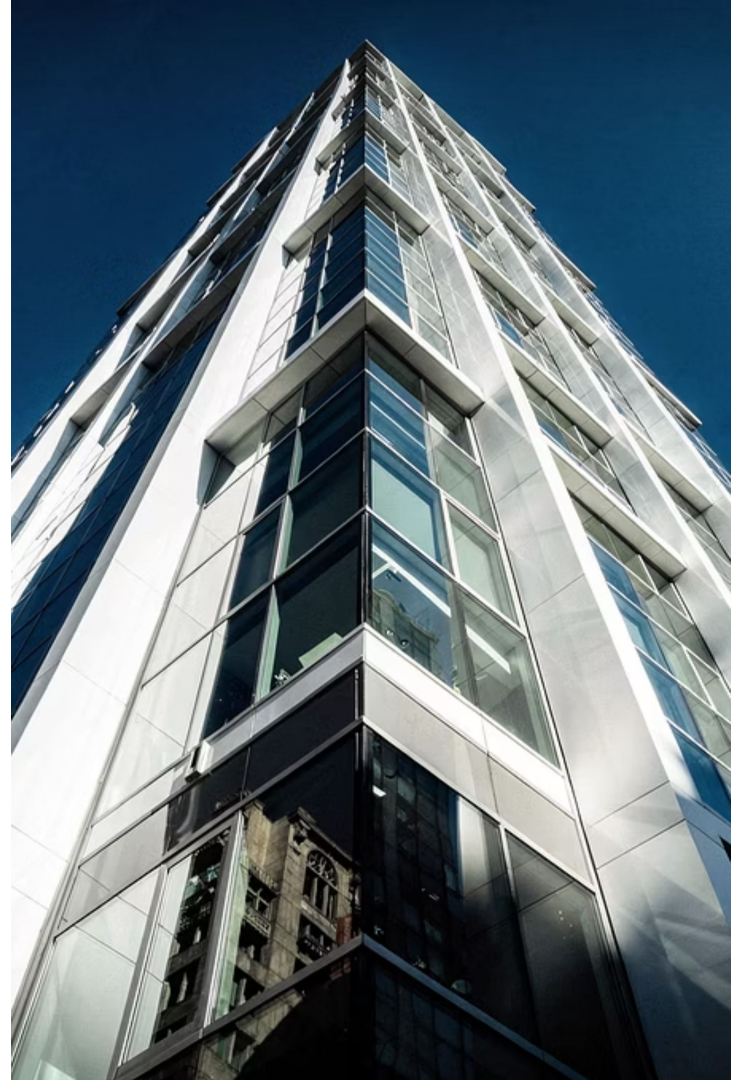
Stay Duration by Hotel Type



Cancellation Risk by Market Segment



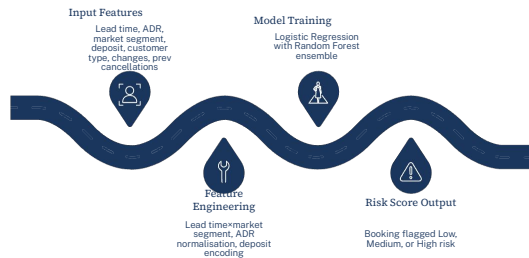
Key Insight: 75% of all bookings are Transient — improving the direct booking experience for this single segment is the highest-impact retention lever available.



# Predicting Cancellations Before They Happen

Logistic Regression + Random Forest ensemble — flags high-risk bookings at reservation time

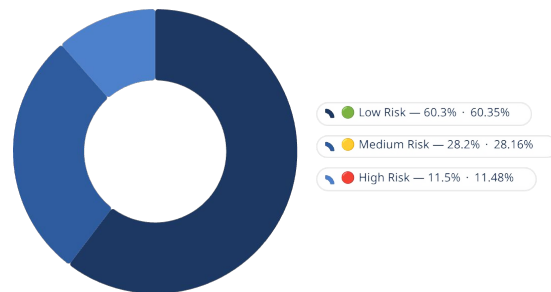
## Model Architecture



## Business Trigger Actions — High Risk Bookings

- **Mandatory Deposit Request**  
Automatically sent at reservation confirmation
- **Reduced Cancellation Window**  
Free cancellation period reduced to 24 hours
- **Prepaid Rate Upsell Offer**  
Convert to non-refundable with incentive discount
- **Revenue Team Alert**  
Manual review flagged for high-value at-risk bookings

## Risk Score Distribution — 119,400 Bookings



72,000  
Low Risk Bookings

33,600  
Medium Risk Bookings

13,700  
High Risk Bookings

13,700 bookings identified as high-risk at point of reservation — enabling proactive revenue protection before cancellation occurs.

# 6 Insights. 3 Pillars. One Clear Direction.

Data-backed findings with action strategies for cancellations, revenue growth, and guest retention

## INSIGHT 01 · RISK

### OTA Dependency Risk

Online TA = highest volume + highest cancellation rate (36.7%). Over-reliance on a single high-risk channel threatens revenue stability.

→ *Prepaid OTA offers · Direct booking campaigns · Stricter cancellation policies*

## INSIGHT 02 · REVENUE

### Peak Season Yield Gap

Jul–Aug peak: ADR reaches \$160 in Room C with volume exceeding 12.7K bookings. Premium pricing is under-exploited during peak demand.

→ *Dynamic pricing model · Surge readiness · Seasonal room packages*

## INSIGHT 03 · REVENUE

### Parking Demand Unmonetised

7,464 parking requests across all segments indicate strong private transport dependence. An untapped ancillary revenue stream.

→ *Premium reserved parking packages · Bundled arrival experience*

## INSIGHT 04 · REVENUE

### Resort Premium Power

Resort ADR exceeds City ADR with stronger pricing power and higher guest spend per stay (4.3 nights vs 3.0 nights).

→ *Luxury packages · Seasonal upsells · Experience bundles*

## INSIGHT 05 · RETENTION

### Transient Dominance

75.1% of bookings come from Transient guests with a 3.2% repeat rate — minimal loyalty investment currently deployed.

→ *Loyalty tier system · Personalised offers · Direct booking incentives*

## INSIGHT 06 · RETENTION

### Channel Over-Dependence

Heavy reliance on OTA platforms for all customer acquisition. Direct channel is underdeveloped, losing margin and guest data.

→ *Website optimisation · Member discounts · Mobile enhancements*

01 · Reduced Cancellation Losses

02 · Increased Revenue & ADR Growth

03 · Improved Guest Retention

# From Data to Decisions – What This Means for Revenue

Three interventions. Measurable impact. Clear next steps.

PROBLEM	DATA EVIDENCE	RECOMMENDED ACTION
<ul style="list-style-type: none"> <li>High OTA Cancellation Risk</li> </ul>	36.7% OTA cancel rate · 44,200 total lost bookings · Groups 61%	Mandatory OTA deposit · Stricter cancellation window · Direct booking campaign
<ul style="list-style-type: none"> <li>Untapped Peak Season Revenue</li> </ul>	ADR peaks \$160 (Room C) · Jul–Aug highest booking volume 12.7K+	Dynamic pricing model · Premium seasonal room packages · Surge-ready inventory
<ul style="list-style-type: none"> <li>Low Guest Retention</li> </ul>	3.2% repeat rate · 75% Transient with zero loyalty investment	Launch loyalty tier program · Personalised direct booking offers · Mobile enhancements

## 12-Month Action Roadmap



Reducing cancellations by just 5% across 119,400 bookings at an ADR of \$101.83 = over \$600,000 in recovered annual revenue.

[Live Power BI Dashboard](#)

[Full PDF Report](#)

[SQL Query](#)

[ML Notebook](#)