HotelRevAI AI Driven Revenue Analysis

Group 3

Dataset Analysis

Brief View of the Dataset

Introduction

The dataset consists of 119,390 rows and 32 columns. It contains booking records from both City Hotels and Resort Hotels, covering information on reservation details, customer demographics, cancellations, revenue (ADR), booking channels, and stay characteristics.

Our goal is to clean, transform, and enrich this dataset so that it can be used to generate key performance indicators (KPIs), trends, and insights for hotel management.

Derived / New Columns

To support KPI calculations, we created the following calculated columns:

1. **TotalNights**
   - **Formula:**
   $$TotalNights = stays\_in\_weekend\_nights + stays\_in\_week\_nights$$
   - **Reason:** Represents total room-nights per booking, which is essential for calculating occupancy and revenue metrics.
2. **RevenuePerBooking**
   - **Formula:**
   $$RevenuePerBooking = adr \times TotalNights$$
   - **Reason:** Directly measures revenue contribution of each booking (excluding cancellations).
3. **ArrivalDate (real date)**
   - **Created from:** *arrival_date_year*, *arrival_date_month*, and *arrival_date_day_of_month*.
   - **Reason:** Simplifies building monthly and weekly trends in Power BI visuals, enabling time-series analysis without complex transformations.

Features Used

1. **hotel**
   - **Why Useful:** Allows comparison of revenue performance between City Hotels and Resort Hotels.
   - **Use in Project:** Analyze and visualize revenue breakdown by hotel type.
2. **is_canceled**
   - **Why Useful:** Required to calculate cancellation rates and identify revenue lost due to cancellations.
   - **Use in Project:** Measure cancellation patterns and assess financial impact.

3. **lead_time**
    - **Why Useful:** Shows how far in advance customers book their stays.
    - **Use in Project:** Supports demand forecasting and booking window analysis.
4. **stays_in_weekend_nights, stays_in_week_nights**
    - **Why Useful:** Represents the number of nights booked.
    - **Use in Project:** Calculate length of stay and derive occupancy and RevPAR metrics.
5. **adults, children, babies**
    - **Why Useful:** Provides guest demographic information.
    - **Use in Project:** Analyze revenue contribution by family size or traveler type.
6. **meal**
    - **Why Useful:** Indicates the meal plan selected (e.g., BB, HB, FB).
    - **Use in Project:** Explore ancillary revenue opportunities through meal preferences.
7. **country**
    - **Why Useful:** Identifies the origin of the guests.
    - **Use in Project:** Build geographic revenue insights using map visualizations.
8. **market_segment, distribution_channel**
    - **Why Useful:** Shows the source of bookings (e.g., Direct, Corporate, OTA).
    - **Use in Project:** Evaluate channel efficiency and dependency on online travel agencies.
9. **is_repeated_guest**
    - **Why Useful:** Identifies loyal or returning customers.
    - **Use in Project:** Track repeat guest rate and loyalty patterns.
10. **reserved_room_type, assigned_room_type**
    - **Why Useful:** Captures room type mismatches due to upgrades or overbooking.
    - **Use in Project:** Analyze operational impacts of mismatches on revenue and customer satisfaction.
11. **deposit_type**
    - **Why Useful:** Indicates payment conditions (No Deposit, Refundable, Non-Refundable).
    - **Use in Project:** Assess cancellation likelihood and financial exposure.
12. **agent**
    - **Why Useful:** Identifies intermediaries responsible for bookings.
    - **Use in Project:** Evaluate top-performing agents and aggregate smaller agents into "Other".
13. **customer_type**
    - **Why Useful:** Differentiates customer categories (Transient, Group, Contract, etc.).
    - **Use in Project:** Segment revenue contribution by customer type.
14. **adr (Average Daily Rate)**
    - **Why Useful:** Core revenue metric indicating the average rate charged per room.
    - **Use in Project:** Compute Total Revenue, ADR trends, and RevPAR.
15. **reservation_status, reservation_status_date**
    - **Why Useful:** Confirms the final outcome of a booking (canceled, checked out, no-show).
    - **Use in Project:** Ensure accurate revenue reporting by excluding invalid bookings.

Conclusion

Through careful cleaning, transformation, and feature selection, the dataset is now structured to directly support hotel revenue intelligence. The following key steps were undertaken:

- We dropped irrelevant or poor-quality features (e.g., company, waiting list).
- We transformed and grouped high-cardinality variables such as agent and fragmented date fields.
- We introduced derived fields including TotalNights, RevenuePerBooking, and a consolidated ArrivalDate.
- We defined a set of KPIs that measure financial health (Revenue, ADR, RevPAR), efficiency (Occupancy, Lead Time), and risks (Cancellations, Revenue Lost).
- We mapped visuals that convert raw data into actionable insights for hotel managers, supporting better strategic and operational decision-making.