Hotel Booking Demand Dataset Documentation

1. Dataset Overview

The **Hotel Booking Demand** dataset contains detailed information about hotel bookings, including customer demographics, booking details, stay duration, and cancellations. The dataset includes information from two types of hotels: **City Hotel (Lisbon, Portugal)** and **Resort Hotel (Algarve, Portugal)**. This dataset is widely used for predictive modeling, customer segmentation, and business intelligence in the hospitality industry.

2. Data Structure

The dataset consists of **119,390 rows** and **32 columns**, including numerical and categorical attributes. The data was extracted directly from the hotels' **Property Management System (PMS) SQL databases** covering bookings from **July 1, 2015, to August 31, 2017**.

2.1 Data Acquisition & Engineering

- Data was acquired using **T-SQL queries** executed directly on the PMS databases.
- Some variables were derived from **multiple PMS tables**.
- The **change log** was used to ensure variables reflected booking status **one day before arrival**.
- Categorical fields like **market segment, meal type, and nationality** were mapped from raw PMS data.

2.2 Numerical Columns

Column Name	Description
lead_time	Number of days between booking and check-in
stays_in_weekend_nights	Number of weekend nights (Saturday & Sunday)
stays_in_week_nights	Number of weeknights (Monday to Friday)
adults	Number of adults per booking
children	Number of children per booking
babies	Number of babies per booking
previous_cancellations	Number of previous cancellations by customer
previous_bookings_not_canceled	Previous successful bookings by the customer
booking_changes	Number of modifications made to the booking

Column Name	Description
days_in_waiting_list	Number of days a booking was on the waiting list
adr	Average Daily Rate (ADR) for the booking

2.3 Categorical Columns

Column Name	Description
hotel	Type of hotel (City Hotel / Resort Hotel)
meal	Type of meal plan (BB, HB, FB, SC)
country	Country of origin of the customer (ISO 3166-3 format)
market_segment	Booking source (Online TA, Offline TA, Corporate, etc.)
distribution_channel	Distribution channel (Direct, TA/TO, GDS)
reserved_room_type	Initially reserved room type
assigned_room_type	Final assigned room type
deposit_type	Type of deposit paid (No Deposit, Non-Refund, etc.)
customer_type	Type of customer (Transient, Contract, Group, etc.)
reservation_status	Status of reservation (Check-Out, Canceled, No-Show)
reservation_status_date	Date when reservation status was updated

3. Data Insights

3.1 Booking Trends

- Most bookings are made for **City Hotels** rather than **Resort Hotels**.
- The **average lead time** for bookings is around **104 days**, with some bookings made more than **700 days** in advance.
- Most customers book stays for **1-3 nights**, with longer stays occurring infrequently.

3.2 Cancellations

- A significant number of bookings get **canceled** before check-in.
- Bookings with **longer lead times** tend to have a **higher cancellation rate**.
- Customers with **previous cancellations** are more likely to cancel again.

3.3 Pricing & Revenue Insights

- The **Average Daily Rate (ADR)** varies widely, with extreme values exceeding **5000**.
- Pricing is higher during **peak seasons** and lower during **off-seasons**.
- Bookings made through **Online Travel Agencies (OTA)** tend to have higher ADRs.

3.4 Customer Demographics

- Most guests come from **Portugal (PRT)**, followed by **United Kingdom (GBR)** and **France (FRA)**.
- Hotels often don't know **customer nationality** until check-in, leading to potential data updates.
- Family bookings with **children and babies** are less common than solo or couple bookings.

3.5 Room Preferences

- Customers often receive **room upgrades or downgrades** based on availability.
- 'Room Type A' is the most frequently booked category.
- Booking attributes like **room type, number of guests, or meal plan** can change at check-in.

4. Business Applications of This Dataset

This dataset is useful for various applications in hospitality analytics, including:

- **Predicting Cancellations:** Machine learning models can be used to forecast whether a booking is likely to be canceled based on lead time, customer history, and other factors.
- **Revenue Optimization:** ADR analysis can help hotels adjust pricing strategies based on seasonal demand.
- **Customer Segmentation:** Identifying key customer groups (business vs. leisure travelers) to tailor marketing efforts.
- **Demand Forecasting:** Understanding booking patterns to optimize hotel occupancy and resource allocation.
- **Enhancing Customer Experience:** Personalizing offerings based on booking trends and room preferences.

5. Conclusion

The **Hotel Booking Demand** dataset provides a rich source of insights for hospitality management, customer analytics, and business intelligence. Understanding the key trends in booking behavior, cancellations, and revenue management can significantly impact hotel performance and customer satisfaction. This dataset can be leveraged for developing **predictive models, business strategies, and customer segmentation frameworks** to optimize hotel operations.

Prepared for Retrieval-Augmented Generation (RAG) Model