Milestone 2 — Dataset Selection & Metadata Description

🏨 **Dataset Name:**

Hotel Booking Demand Dataset

📍 **Source:**

Kaggle (publicly available dataset)

Originally published by **Antonio et al.** for hospitality demand research

---

📊 **Dataset Overview**

This dataset contains booking information from two types of hotels:

HOTEL TYPE	DESCRIPTION
CITY HOTEL	Located in an urban city area, suited for business travel
RESORT HOTEL	Located in a leisure/tourism area, suited for vacations

The dataset includes details on booking behavior, guest composition, pricing, cancellations, and customer segmentation.

📑 **Dataset Size**

METRIC	VALUE
ORIGINAL ROWS	119,390 bookings
AFTER CLEANING	117,399 bookings
COLUMNS	32 attributes

📃 **Metadata (Key Columns Description)**

COLUMN NAME	TYPE	DESCRIPTION
HOTEL	Categorical	Hotel category: City or Resort
IS_CANCELED	Binary (0/1)	1 = Booking canceled, 0 = Successful stay
LEAD_TIME	Integer	Days between booking date and arrival
ARRIVAL_DATE_YEAR	Integer	Year of arrival
ARRIVAL_DATE_MONTH	Categorical	Month of arrival
ARRIVAL_DATE_WEEK_NUMBER	Integer	Week number of arrival
ARRIVAL_DATE_DAY_OF_MONTH	Integer	Day of arrival
STAYS_IN_WEEKEND_NIGHTS	Integer	Weekend nights stayed
STAYS_IN_WEEK_NIGHTS	Integer	Weekday nights stayed
ADULTS	Integer	Number of adults in booking
CHILDREN	Integer	Number of children
BABIES	Integer	Number of babies

MEAL	Categorical	Type of meal plan booked (BB, HB, SC...)
COUNTRY	Categorical	Guest country code
MARKET_SEGMENT	Categorical	Segment of booking (Direct, Online, Corporate...)
DISTRIBUTION_CHANNEL	Categorical	Booking source (TA/OTA, Direct...)
IS_REPEATED_GUEST	Binary	1 = Returning guest
PREVIOUS_CANCELLATIONS	Integer	Number of past canceled bookings
RESERVED_ROOM_TYPE	Categorical	Initially reserved room type
ASSIGNED_ROOM_TYPE	Categorical	Actual room assigned at check-in
BOOKING_CHANGES	Integer	Number of modifications made to booking
DEPOSIT_TYPE	Categorical	Deposit status: No Deposit / Non-Refund / Refundable
AGENT	Numeric	Travel agent ID
COMPANY	Numeric	Company that booked
DAYS_IN_WAITING_LIST	Integer	Waiting time for confirmation
CUSTOMER_TYPE	Categorical	Type of guest (Transient, Group, Contract...)
ADR	Float	Average Daily Rate (Revenue per room per night)
REQUIRED_CAR_PARKING_SPACES	Integer	Parking spaces requested
TOTAL_OF_SPECIAL_REQUESTS	Integer	Additional requests by guest
RESERVATION_STATUS	Categorical	Check-Out / Canceled / No-Show
RESERVATION_STATUS_DATE	Date	Status update date

🎯 Relevance of the Dataset to Project KPIs

KPI	COLUMNS USED
TOTAL REVENUE	adr × stays columns
OCCUPANCY RATE	booking vs cancellations
REVPAR	Revenue per booking
SEASONAL TRENDS	arrival date columns
CUSTOMER SEGMENTATION	customer_type
CHANNEL PERFORMANCE	market_segment & distribution_channel

⚙ Data Cleaning Performed (Summary)

ACTION	REASON
REMOVED BOOKINGS WITH 0 GUESTS	Invalid records
REMOVED BOOKINGS WITH ADR = 0	Free rooms / errors
FILLED MISSING VALUES IN AGENT & COMPANY	Maintain consistency

RETAINED CHILDREN/COUNTRY MISSING WITH CAUTION	Minimal loss
CREATED NEW COLUMNS: TOTAL NIGHTS, MONTH NUMBER, ARRIVALDATE	For analysis and dashboard visuals

✔ **Milestone 2 Conclusion**

This dataset provides comprehensive booking and revenue data required for analyzing:

- Customer behavior
- Seasonal revenue patterns
- Booking channel performance
- Cancellation impacts
- Revenue optimization insights