Low-Level Design (LLD) – Hotel Booking Performance Analysis

Domain: Travel & Hospitality
Assessment Type: Milestone
Difficulty: Medium–High | Total Marks: 20
Concepts Covered:

- DataFrame creation

- Multi-column grouping

- Date parsing and transformation

- Filtering

- Missing value handling

- Data summarization and formatting

Problem Statement

You are working with a travel agency analyzing hotel booking logs to study booking behaviors, cancellations, and hotel performance. Each record includes guest name, hotel name, booking date, status (Booked / Cancelled / No-show), and booking amount. Your goal is to extract insights, spot trends, and clean bad data using Pandas.

Functions to Implement

---

1. Create Booking DataFrame

def create_booking_df(booking_data: list) -> pd.DataFrame:

Input:

[

  ["Alice", "Hotel Blue", "2024-06-01", "Booked", 300.0],

  ["Bob", "Hotel Red", "2024-06-02", "Cancelled", 0.0]

]

Output Columns:
Guest, Hotel, Date, Status, Amount

---

2. Monthly Revenue Per Hotel

python

CopyEdit

```python
def compute_monthly_revenue(df: pd.DataFrame) -> pd.DataFrame:
```

Logic:

- Extract Month (YYYY-MM) from Date

- Only consider Booked records

- Group by Hotel + Month and sum Amount

---

3. Add Cancellation Flag

python

CopyEdit

```python
def add_cancellation_flag(df: pd.DataFrame) -> pd.DataFrame:
```

Output: Add a column IsCancelled = 1 if Status == "Cancelled" else 0

---

4. Frequent No-Show Guests

```python
def get_no_show_guests(df: pd.DataFrame, threshold: int) -> pd.DataFrame:
```

Logic:

- Group by Guest

- Count where Status == "No-show"

- Return those with count > threshold

---

5. Hotel Booking Status Summary

```python
def hotel_status_summary(df: pd.DataFrame) -> pd.DataFrame:
```

Logic:

- Group by Hotel and Status

- Use .unstack(fill_value=0)

- Return counts of each status per hotel

---

6. Clean Booking Data

```python
def clean_booking_data(df: pd.DataFrame) -> pd.DataFrame:
```

Logic:

- Drop rows where Status is not in ["Booked", "Cancelled", "No-show"]

- Drop rows with null Amount

Test Case Matrix

TC ID	Description	Function	Marks
TC1	Create structured booking DataFrame	create_booking_df()	2.5
TC2	Compute monthly revenue per hotel	compute_monthly_revenue()	2.5
TC3	Add cancellation flag column	add_cancellation_flag()	2.5
TC4	Identify guests with frequent no-shows	get_no_show_guests()	2.5
TC5	Hotel-wise status summary	hotel_status_summary()	2.5
HTC1	Clean records with null/invalid status	clean_booking_data()	2.5
HTC2	Handle guests with only booked stays	compute_monthly_revenue()	2.5
HTC3	Hotel with only one type of status	hotel_status_summary()	2.5

Example Input & Output (TC1)

create_booking_df([

  ["Alice", "Hotel Blue", "2024-06-01", "Booked", 300.0],

  ["Bob", "Hotel Red", "2024-06-02", "Cancelled", 0.0]

])

Expected Output:

Guest	Hotel	Date	Status	Amount
Alice	Hotel Blue	2024-06-01	Booked	300.0
Bob	Hotel Red	2024-06-02	Cancelled	0.0