Hotel Reservations Report

1. Deliverables
    - Python Code: W3P2.ipynb
    - SQL text file (.sql): SQL file hotels
    - Short Presentation:
    - PDF Document - this document

2. Data Dictionary

Booking	Booking Details
• Booking_ID: unique identifier of each booking • no_of_adults: Number of adults • no_of_children: Number of Children • no_of_weekend_nights: Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel • no_of_week_nights: Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel • type_of_meal_plan: Type of meal plan booked by the customer: • required_car_parking_space: Does the customer require a car parking space? (0 - No, 1- Yes) • room_type_reserved: Type of room reserved by the customer. The values are ciphered (encoded) by INN Hotels. • lead_time: Number of days between the date of booking and the arrival date • arrival_year: Year of arrival date • arrival_month: Month of arrival date • arrival_date: Date of the month	• Booking_ID: unique identifier of each booking • market_segment_type: Market segment designation. • repeated_guest: Is the customer a repeated guest? (0 - No, 1- Yes) • no_of_previous_cancellations: Number of previous bookings that were cancelled by the customer prior to the current booking • no_of_previous_bookings_not_canceled: Number of previous bookings not cancelled by the customer prior to the current booking • avg_price_per_room: Average price per day of the reservation; prices of the rooms are dynamic. (in euros) • no_of_special_requests: Total number of special requests made by the customer (e.g. high floor, view from the room, etc) • booking_status: Flag indicating if the booking was cancelled or not.

3. Python Documentation

Exploratory Data Analysis (EDA)

1. Data Import and Initial Inspection:
    - The dataset was imported using pandas and initial inspection was conducted with df.head() and df.describe().
    - Column names were printed for reference.
2. Data Cleaning:
    - Rows with missing values were removed using df.dropna(inplace=True).
    - Duplicated rows were also removed using df.drop_duplicates(inplace=True).
3. Consistency Checks:
    - For categorical columns (e.g., booking_status), unique values were identified to check for inconsistencies.
    - For numerical columns (e.g., no_of_adults, no_of_children), checks for negative and non-integer values were performed to ensure data integrity.
4. Distribution Analysis:
    - Histograms for numerical columns were plotted to analyse their distributions.
    - Bar plots for categorical columns (e.g., type_of_meal_plan, market_segment_type, booking_status) were created to understand their frequencies.
    - The distribution of the target variable booking_status was specifically examined.
5. Feature Selection:
    - Two sets of columns were defined for separate analysis:
        - columns_table1: Contains demographic and booking detail columns.
        - columns_table2: Contains market segment information and booking status details.
6. DataFrame Creation:
    - Separate DataFrames (booking and bookingdetails) were created using the specified column sets for more focused analysis.

Additional Analyses

1. Booking Trends Analysis
2. Guest Preferences
3. Cancellation Analysis
4. Market Segment Analysis
5. Repeat Guest Behavior
6. Weekend vs. Weeknight Stays
7. Seasonal Booking Trends

8. Impact of Special Requests
9. Booking Behavior of Families
10. Parking Space Requirement Analysis

Web Scraping

- Hotel Médano Website:
    - Used requests and BeautifulSoup to scrape the hotel's website.
    - Extracted specific textual information from div elements containing class containers.

Data Export to MySQL

1. Database Connection:
    - Connected to a MySQL database using SQLAlchemy.
    - Defined MySQL connection parameters and created an engine for database interaction.
2. Data Export:
    - Exported the booking and bookingdetails DataFrames to the MySQL database.
    - Ensured the data was stored in a specified schema and confirmed successful data transfer.

Questions:

1. How long do people with kids stay vs. people with no kids?
1. Which segments order meal plans?
2. How do the number of bookings and cancellations vary by year and month? What are the seasonal trends in bookings?
3. What are the different segments of customers based on their booking behaviours, such as repeated guests versus new guests?
4. Which customer segments have the highest cancellation rates?
5. How does the average price per room vary by room type and season?
6. Market Segment Performance:
7. How do different market segments (e.g., leisure, business) perform in terms of number of bookings, average price per room, and lead time?
8. How do special requests impact the likelihood of cancellation?
9. What are the most popular room types among different customer segments?