**Project:- Hotel Booking Analysis using SQL**

**Table of Contents**

1. Introduction
2. Dataset Description
3. Data Cleaning Steps
4. Database & Table Creation
5. Data Import Process
6. Exploratory Data Analysis (EDA) Queries
7. Key Findings
8. Conclusion
9. Future Work

**1. Introduction**

This project focuses on analyzing hotel booking data using SQL. The aim is to extract valuable insights such as peak booking periods, customer preferences, cancellation trends, and other operational metrics that can support decision-making.

**2. Dataset Description**

The dataset includes various fields such as hotel type, booking dates, number of guests, assigned room types, market segment, booking status, and customer information. It contains the following key columns:

* hotel, is\_canceled, lead\_time, arrival\_date\_year, arrival\_date\_month, arrival\_date\_week\_number, arrival\_date\_day\_of\_month
* stays\_in\_weekend\_nights, stays\_in\_week\_nights, adults, children, babies
* meal, country, market\_segment, distribution\_channel
* agent, company, days\_in\_waiting\_list, customer\_type
* adr (average daily rate), required\_car\_parking\_spaces, total\_of\_special\_requests
* reservation\_status, reservation\_status\_date
* name, email, phonenumber, credit\_card

**3. Data Cleaning Steps**

* **Duplicate Removal**:

WITH duplicate\_rows AS (

SELECT MIN(id) AS min\_id

FROM hotel\_booking

GROUP BY hotel, arrival\_date\_year, arrival\_date\_month, arrival\_date\_day\_of\_month, adr

)

DELETE FROM hotel\_booking

WHERE id NOT IN (SELECT min\_id FROM duplicate\_rows);

* **Handling NULL Values**:

UPDATE hotel\_booking SET children = 0 WHERE children IS NULL;

UPDATE hotel\_booking SET babies = 0 WHERE babies IS NULL;

UPDATE hotel\_booking SET agent = 0 WHERE agent IS NULL;

UPDATE hotel\_booking SET company = 0 WHERE company IS NULL;

UPDATE hotel\_booking SET country = 'Unknown' WHERE country IS NULL;

* **Other Cleaning**:

DELETE FROM hotel\_booking WHERE adults = 0 AND children = 0 AND babies = 0;

UPDATE hotel\_booking SET lead\_time = ABS(lead\_time) WHERE lead\_time < 0;

ALTER TABLE hotel\_booking ADD total\_stay INT;

UPDATE hotel\_booking SET total\_stay = stays\_in\_weekend\_nights + stays\_in\_week\_nights;

**4. Database & Table Creation**

CREATE TABLE hotel\_booking (

id INT PRIMARY KEY AUTO\_INCREMENT,

hotel VARCHAR(50),

is\_canceled TINYINT,

lead\_time INT,

arrival\_date\_year INT,

arrival\_date\_month VARCHAR(15),

arrival\_date\_week\_number INT,

arrival\_date\_day\_of\_month INT,

stays\_in\_weekend\_nights INT,

stays\_in\_week\_nights INT,

adults INT,

children INT,

babies INT,

meal VARCHAR(10),

country VARCHAR(5),

market\_segment VARCHAR(50),

distribution\_channel VARCHAR(50),

is\_repeated\_guest TINYINT,

previous\_cancellations INT,

previous\_bookings\_not\_canceled INT,

reserved\_room\_type VARCHAR(5),

assigned\_room\_type VARCHAR(5),

booking\_changes INT,

deposit\_type VARCHAR(20),

agent INT,

company INT,

days\_in\_waiting\_list INT,

customer\_type VARCHAR(20),

adr FLOAT,

required\_car\_parking\_spaces INT,

total\_of\_special\_requests INT,

reservation\_status VARCHAR(20),

reservation\_status\_date DATE,

name VARCHAR(20),

email VARCHAR(30),

phonenumber VARCHAR(20),

credit\_card VARCHAR(20)

);

**5. Data Import Process**

* Imported using MySQL Workbench or similar tools
* Faced issues with incorrect datatypes (e.g. empty strings in integer columns)
* Resolved by modifying column types and handling blanks/nulls properly

**6. Exploratory Data Analysis (EDA) Queries**

**Example Queries with Explanation:**

* **Total Number of Bookings:**

SELECT COUNT(\*) AS total\_bookings FROM hotel\_booking;

Shows how many bookings are in the dataset.

* **Peak Booking Months:**

SELECT arrival\_date\_month, COUNT(\*) AS total\_bookings

FROM hotel\_booking

GROUP BY arrival\_date\_month

ORDER BY total\_bookings DESC;

Identifies months with the most bookings.

* **Most Common Meal Type:**

SELECT meal, COUNT(\*) AS total\_count

FROM hotel\_booking

GROUP BY meal

ORDER BY total\_count DESC;

Reveals the preferred meal plan.

*(More than 25 total queries are included in the hotel booking SQL section)*

**7. Key Findings**

* City hotel receives more bookings than resort hotel
* Most bookings happen in August and July
* Room type A is most commonly reserved
* Majority of guests do not cancel their bookings
* Most customers are transient (non-repeated)
* The adr (average daily rate) is higher for city hotels

**8. Conclusion**

This SQL-based project helped in exploring real-world data handling and analysis. It gave practical exposure to data cleaning, complex querying, and drawing insights using SQL.