Case Study by A Major Travel Company

SQL FOR DATA ANALYTICS

- UNDERSTANDING THE PROBLEM STATEMENT
- SOLVING BUSINESS PROBLEM USING 4 QUESTIONS

DATASET :

```
45 •  Select * from user_table ut inner join booking_table bt
46    on ut.user_id =bt.User_id;
47
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

User_id	Segment	Booking_id	Booking_date	User_id	Line_of_business
u1	s1	b1	2022-03-23	u1	Flight
u2	s1	b2	2022-03-27	u2	Flight
u1	s1	b3	2022-03-28	u1	Hotel
u4	s2	b4	2022-03-31	u4	Flight
u1	s1	b5	2022-04-02	u1	Hotel
u2	s1	b6	2022-04-02	u2	Flight
u5	s2	b7	2022-04-06	u5	Flight
u6	s3	b8	2022-04-06	u6	Hotel
u2	s1	b9	2022-04-06	u2	Flight
u1	s1	b10	2022-04-10	u1	Flight

QS .1 WRITE A SQL QUERY WHICH GIVES SEGMENT WISE TOTAL USERS WHO BOOKED FLIGHT IN APRIL 2022 ?

```
48    /*Qs.1 Write a SQL query that gives the below output ( Summary at Segment Level- Give Segment wise Total Users who booked Flight in April 2022)*/
49
50 •  Select ut.Segment, count(distinct(ut.User_id)) as Total_user_count,
51    count(distinct case when bt.Line_of_business = "Flight" and bt.Booking_date between '2022-04-01' and '2022-04-30'then bt.User_id end) as Users_who_booked_
52    from user_table ut  left join booking_table bt
53    on ut.User_id = bt.User_id
54    group by segment;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Segment	Total_user_count	Users_who_booked_flight_in_Apr2022
s1	3	2
s2	2	2
s3	5	1

Result Grid

Form Editor

QS.2 WRITE A QUERY TO IDENTIFY USERS WHOSE FIRST BOOKING WAS A “HOTEL” BOOKING

```
56    /*Qs.2 Write a query to identify users whose first booking was a Hotel booking*/
57
58    With hotel_booking as (
59    Select * ,
60    rank() over (partition by user_id order by booking_date) as rn
61    from booking_table)
62    Select * from hotel_booking
63    where rn =1 and Line_of_business ="Hotel";
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Booking_id	Booking_date	User_id	Line_of_business	rn
b8	2022-04-06	u6	Hotel	1

QS.3.WRITE A QUERY TO CALCULATE THE DAYS BETWEEN FIRST AND LAST BOOKING OF EACH USER

```
65    /*Qs.3 Write a query to calculate the days between first  and last booking of each user*/
66
67    Select User_id, max(Booking_date), min(Booking_date), datediff(max(booking_date), min(booking_date)) as Days_between_First_and_Last_Booking
68    from booking_table
69    group by user_id;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

User_id	max(Booking_date)	min(Booking_date)	Days_between_First_and_Last_Booking
u1	2022-05-06	2022-03-23	44
u2	2022-04-28	2022-03-27	32
u4	2022-05-04	2022-03-31	34
u5	2022-04-20	2022-04-06	14
u6	2022-04-22	2022-04-06	16

QS.4 WRITE A QUERY TO COUNT THE NUMBER OF FLIGHT AND HOTEL BOOKINGS IN EACH OF THE USER SEGMENTS FOR THE YEAR 2022

You can filter by year since only 2022 in dataset is given have not used filter in this scenario

```
70    /* Write a query to count the number of flight and hotel bookings in each of the user segments for the year 2022*/
71
72 •  Select Segment,
73    sum(case when Line_of_business ='Flight' then 1 else 0 end )as flight_booking,
74    sum(case when Line_of_business ='Hotel' then 1 else 0 end) as hotel_booking
75    from booking_table bt join user_table ut
76    on ut.User_id = bt.User_id
77    group by Segment;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Segment	flight_booking	hotel_booking
s1	8	4
s2	3	3
s3	1	1