SQL Capstone Project

Please answer the following questions using Airline DB database.

Instruction to attempt questions:

- Students need to write queries for the questions mentioned in the using Airline DB database
- Read the questions carefully before writing the query in **Airline Playground** (in the Playground chapter of SQL)
- Airline DB: https://www.skillovilla.com/playground/sql?exerciseId=0181e251-6ea8-4595-ae2b-0c690119f8db

How to submit the capstone:

- Copy the SQL query code and paste it in the answer section in this file.
- Once the assignment is done, submit the file over LMS.

Invalid Submissions:

- Pasting pictures of the code as answer is **NOT** acceptable.
- Uploading output data (CSVs) of the SQL queries is **NOT** acceptable.

Write your answers(query) in the answer and submit it. To write the answer in the assignment, please follow the below example in yellow

Example:

Questions*: Extract all the columns of the flights table*

Answer: *SELECT * FROM flights*

Attempt the following Questions-

1. ***Represent the "book_date" column in "yyyy-mmm-dd" format using Bookings table***

Expected output: book_ref, book_date (in "yyyy-mmm-dd" format) , total amount

Answer:

```
SELECT
    book_ref,
    TO_CHAR( book_date , 'yyyy-mon-dd'),
    total_amount
FROM bookings;
```

2. **Get the following columns in the exact same sequence.**

Expected columns in the output: ticket_no, boarding_no, seat_number, passenger_id, passenger_name.

Answer:

```
SELECT
    t.ticket_no,
    b.boarding_no,
    b.seat_no,
    t.passenger_id,
    t.passenger_name
FROM tickets t
JOIN boarding_passes b
ON t.ticket_no = b.ticket_no;
```

3. Write a query to find the seat number which is least allocated among all the seats?

Answer:

```
WITH seat_allocation_rankings AS
(
    SELECT
        s.seat_no,
        COUNT(*) AS count_of_seat_allocation,
        DENSE_RANK()
        OVER( ORDER BY COUNT(*) ) AS allocation_rankings
    FROM seats s
    JOIN boarding_passes b
    ON s.seat_no = b.seat_no
    GROUP BY 1
)

SELECT seat_no
FROM seat_allocation_rankings
WHERE allocation_rankings = 1
;
```

4. ***In the database, identify the month wise highest paying passenger name and passenger id.***

Expected output: Month_name("mmm-yy" format), passenger_id, passenger_name and total amount

Answer:

```
WITH monthly_highest_paying_passenger_rankings AS
(
   SELECT
      TO_CHAR(b.book_date , 'mon-yy') AS month_name,
      t.passenger_id,
      t.passenger_name,
      SUM(b.total_amount) AS total_amount,
      DENSE_RANK()
      OVER( PARTITION BY TO_CHAR(b.book_date , 'mon-yy')
ORDER BY SUM(b.total_amount) DESC ) AS
passenger_payment_rankings
   FROM bookings b
   JOIN tickets t
   ON b.book_ref = t.book_ref
   GROUP BY 1 ,2 ,3
)

SELECT
   month_name,
   passenger_id,
   passenger_name,
   total_amount
FROM monthly_highest_paying_passenger_rankings
WHERE passenger_payment_rankings = 1
;
```

5. ***In the database, identify the month wise least paying passenger name and passenger id?***

Expected output: Month_name(“mmm-yy” format), passenger_id, passenger_name and total amount

Answer:

```
WITH monthly_least_paying_passenger_rankings AS
(
   SELECT
      TO_CHAR(b.book_date , 'mon-yy') AS month_name,
      t.passenger_id,
      t.passenger_name,
      SUM(b.total_amount) AS total_amount,
      DENSE_RANK()
      OVER( PARTITION BY TO_CHAR(b.book_date , 'mon-yy')
ORDER BY SUM(b.total_amount) ) AS
passenger_payment_rankings
   FROM bookings b
   JOIN tickets t
   ON b.book_ref = t.book_ref
   GROUP BY 1 ,2 ,3
)

SELECT
   month_name,
   passenger_id,
   passenger_name,
   total_amount
FROM monthly_highest_paying_passenger_rankings
WHERE passenger_payment_rankings = 1
;
```

6. **Identify the travel details of non stop journeys or return journeys (having more than 1 flight).**

Expected Output: Passenger_id, passenger_name, ticket_number and flight count.

Answer:

```
SELECT
   t.passenger_id,
   t.passenger_name,
   t.ticket_no,
   COUNT(t_f.flight_id) AS flight_count
FROM ticket_flights t_f
JOIN tickets t
ON t_f.ticket_no = t.ticket_no
GROUP BY 1 , 2 , 3
HAVING COUNT(t_f.flight_id) > 1;
```

7. **How many tickets are there without boarding passes?**
Expected Output: just one number is required.

Answer:

```
SELECT
   COUNT(t.ticket_no)
FROM tickets t
LEFT JOIN boarding_passes b
ON t.ticket_no = b.ticket_no
WHERE b.ticket_no IS NULL
;
```

8. **Identify details of the longest flight (using flights table)?**
Expected Output: Flight number, departure airport, arrival airport, aircraft code and durations.

Answer:

```
WITH flight_duration_rankings AS
   (
      SELECT
         flight_no,
         departure_airport,
         arrival_airport,
         aircraft_code,
         (scheduled_arrival - scheduled_departure) AS
durations,
         DENSE_RANK()
         OVER( ORDER BY (scheduled_arrival -
scheduled_departure) DESC ) AS duration_rankings
   FROM flights
   GROUP BY 1 ,2 ,3 ,4 , scheduled_arrival ,
scheduled_departure
   )

   SELECT
      flight_no,
      departure_airport,
      arrival_airport,
      aircraft_code,
      durations
   FROM flight_duration_rankings
   WHERE duration_rankings = 1;
```

9. **Identify details of all the morning flights (morning means between 6AM to 11 AM, using flights table)?**

Expected output: flight_id, flight_number, scheduled_departure, scheduled_arrival and timings.

Answer:

```
SELECT
    flight_id,
    flight_no,
    scheduled_departure,
    scheduled_arrival,
    scheduled_departure::time AS departure_time,
    scheduled_arrival::time AS arrival_time
FROM flights
WHERE (scheduled_departure::time BETWEEN
'06:00:00.000+0000' AND '11:00:00.000+0000')
AND (scheduled_arrival::time BETWEEN '06:00:00.000+0000'
AND '11:00:00.000+0000');
```

10. **Identify the earliest morning flight available from every airport.**

Expected output: flight_id, flight_number, scheduled_departure, scheduled_arrival, departure airport and timings.

Answer:

```
WITH airlines_arrival_rankings AS
(
   SELECT
     flight_id,
     flight_no,
     scheduled_departure,
     scheduled_arrival,
     departure_airport,
     scheduled_departure::time AS departure_time,
     scheduled_arrival::time AS arrival_time,
     DENSE_RANK()
     OVER( PARTITION BY departure_airport ORDER BY
scheduled_departure::time ) AS timmings_rankings
   FROM flights
   WHERE (scheduled_departure::time BETWEEN '06:00:00.000+0000'
AND '11:00:00.000+0000')
   AND (scheduled_arrival::time BETWEEN '06:00:00.000+0000' AND
'11:00:00.000+0000')

)

SELECT
  flight_id,
  flight_no,
  scheduled_departure,
  scheduled_arrival,
  departure_airport,
  (departure_time || ' to ' || arrival_time) AS timings
FROM airlines_arrival_rankings
WHERE timmings_rankings = 1;
```

11. **Questions: Find list of airport codes in Europe/Moscow timezone**
Expected Output: Airport_code.

Answer:

```
SELECT
   airport_code
FROM airports
WHERE timezone = 'Europe/Moscow';
```

12. **Write a query to get the count of seats in various fare condition for every aircraft code?**
Expected Outputs: Aircraft_code, fare_conditions ,seat count

Answer:
```
SELECT
   aircraft_code,
   fare_conditions,
   COUNT(seat_no) AS seat_count
FROM seats
GROUP BY 1,2
;
```

13. How many aircrafts codes have at least one Business class seats?
Expected Output : Count of aircraft codes

Answer:
```
SELECT
   COUNT(aircraft_code) AS count_of_aircraft_codes
FROM seats
WHERE fare_conditions = 'Business'
;
```

14. Find out the name of the airport having maximum number of departure flight
Expected Output : Airport_name

Answer:
```
WITH departure_flights_rankings AS
(
   SELECT
      a.airport_name,
      DENSE_RANK()
      OVER( ORDER BY COUNT(*) DESC ) AS flights_rankings
   FROM airports a
   JOIN flights f
   ON a.airport_code = f.departure_airport
   GROUP BY 1
)

SELECT
   airport_name
FROM departure_flights_rankings
WHERE flights_rankings = 1;
```

15. Find out the name of the airport having least number of scheduled departure flights
Expected Output : Airport_name

Answer:
```
WITH departure_flights_rankings AS
(
   SELECT
      a.airport_name,
      DENSE_RANK()
      OVER( ORDER BY COUNT(*) ) AS flights_rankings
   FROM airports a
   JOIN flights f
   ON a.airport_code = f.departure_airport
   GROUP BY 1
)

SELECT
   airport_name
FROM departure_flights_rankings
WHERE flights_rankings = 1;
```

16. How many flights from 'DME' airport don't have actual departure?
Expected Output : Flight Count

Answer:
```
SELECT
   COUNT(flight_no)
FROM flights
WHERE departure_airport = 'DME'
AND actual_departure IS NULL;
```

17. Identify flight ids having range between 3000 to 6000
Expected Output : Flight_Number , aircraft_code, ranges

Answer:
```
SELECT
    f.flight_no,
    ar.aircraft_code,
    ar.range
FROM flights f
JOIN aircrafts ar
ON f.aircraft_code = ar.aircraft_code
WHERE range BETWEEN 3000 AND 6000;
```

18. **Write a query to get the count of flights flying between URS and KUF?**
Expected Output : Flight_count

Answer:
```
SELECT
    COUNT(flight_id) AS flight_count
FROM flights
WHERE (departure_airport = 'URS'
AND arrival_airport = 'KUF')
OR
(arrival_airport = 'URS'
AND departure_airport = 'KUF');
```

19. **Write a query to get the count of flights flying from either from NOZ or KRR?**
Expected Output : Flight count

Answer:
```
SELECT
    COUNT(flight_id) AS flight_count
FROM flights
WHERE departure_airport = 'NOZ'
OR departure_airport = 'KRR'
;
```

20. **Write a query to get the count of flights flying from KZN,DME,NBC,NJC,GDX,SGC,VKO,ROV**
Expected Output : Departure airport ,count of flights flying from these airports.

Answer:
```
SELECT
    departure_airport,
    COUNT(flight_id) AS flight_count
FROM flights
WHERE departure_airport IN
('KZN','DME','NBC','NJC','GDX','SGC','VKO','ROV')
GROUP BY 1
;
```

21. **Write a query to extract flight details having range between 3000 and 6000 and flying from DME**

Expected Output :Flight_no , aircraft_code , range , departure_airport

Answer:

```
SELECT
    f.flight_no,
    f.aircraft_code,
    ar.range,
    f.departure_airport
FROM flights f
JOIN aircrafts ar
ON f.aircraft_code = ar.aircraft_code
WHERE departure_airport = 'DME'
AND ar.range BETWEEN 3000 AND 6000;
```

22. **Find the list of flight ids which are using aircrafts from “Airbus” company and got cancelled or delayed**

Expected Output : Flight_id , aircraft_model

Answer:

```
SELECT
    f.flight_id,
    ar.model
FROM flights f
JOIN aircrafts ar
ON f.aircraft_code = ar.aircraft_code
WHERE ar.model LIKE '%Airbus%'
AND (f.status = 'Cancelled' OR f.status = 'Delayed');
```

23. **Find the list of flight ids which are using aircrafts from “Boeing” company and got cancelled or delayed**

Expected Output : Flight_id , aircraft_model

Answer:

```
SELECT
    f.flight_id,
    ar.model
FROM flights f
JOIN aircrafts ar
ON f.aircraft_code = ar.aircraft_code
WHERE ar.model LIKE '%Boeing%'
AND (f.status = 'Cancelled' OR f.status = 'Delayed');
```

24. **Which airport(name) has most cancelled flights (arriving)?**

Expected Output : Airport_name

Answer:

```
WITH flight_rankings AS
(SELECT
   a.airport_name,
   COUNT(*),
   DENSE_RANK()
   OVER( ORDER BY COUNT(*) DESC ) AS cancelled_flights_rankings
FROM airports a
JOIN flights f
ON a.airport_code = f.arrival_airport
WHERE f.status = 'Cancelled'
GROUP BY 1)

SELECT
   airport_name
FROM flight_rankings
WHERE cancelled_flights_rankings = 1;
```

25. Identify flight ids which are using "Airbus aircrafts"
Expected Output : Flight_id , aircraft_model

Answer:

```
SELECT
   f.flight_id,
   ar.model AS aircraft_model
FROM flights f
JOIN aircrafts ar
ON f.aircraft_code = ar.aircraft_code
WHERE model LIKE '%Airbus%';
```

26. Identify date-wise last flight id flying from every airport?
Expected Output: Flight_id , flight_number , schedule_departure , departure_airport

Answer:
```
WITH date_wise_flight_rankings AS
(
   SELECT
      flight_id,
      flight_no,
      scheduled_departure,
      departure_airport,
      DENSE_RANK()
      OVER( PARTITION BY departure_airport ORDER BY
scheduled_departure DESC ) AS flight_rankings
   FROM flights
)

SELECT
   flight_id,
   flight_no,
   scheduled_departure,
   departure_airport
FROM date_wise_flight_rankings
WHERE flight_rankings = 1;
```

27. ***Identify list of customers who will get the refund due to cancellation of the flights and how much amount they will get?***

Expected Output : Passenger_name , total_refund.

Answer:
```
SELECT
   t.passenger_name,
   SUM(tf.amount) AS total_refund
FROM
flights f
JOIN
 ticket_flights tf ON f.flight_id = tf.flight_id
JOIN
 tickets t ON tf.ticket_no = t.ticket_no
WHERE
 f.status = 'Cancelled'
GROUP BY 1;
```

28. ***Identify date wise first cancelled flight id flying for every airport?***

Expected Output : Flight_id , flight_number , schedule_departure ,departure_airport

Answer:
```
WITH cancelled_out_flights_rankings AS
(
   SELECT
      flight_id,
      flight_no,
      scheduled_departure,
      departure_airport,
      DENSE_RANK()
      OVER( PARTITION BY departure_airport ORDER BY
scheduled_departure ) AS cancelled_flights_rankings
   FROM flights
   WHERE status = 'Cancelled'
)

SELECT
   flight_id,
   flight_no,
   scheduled_departure,
   departure_airport
FROM cancelled_out_flights_rankings
WHERE cancelled_flights_rankings = 1;
```

29. Identify list of Airbus flight ids which got cancelled.
Expected Output : Flight_id

Answer:
```
SELECT
   f.flight_id
FROM flights f
JOIN aircrafts ar
ON f.aircraft_code = ar.aircraft_code
WHERE f.status = 'Cancelled'
AND ar.model = '%Airbus%';
```

30. Identify list of flight ids having highest range.
Expected Output : Flight_no, range

Answer:

```
WITH flights_range_rankings AS
(
   SELECT
     f.flight_no,
     ar.range,
     DENSE_RANK()
     OVER( ORDER BY ar.range DESC ) AS range_rankings
   FROM flights f
   JOIN aircrafts ar
   ON f.aircraft_code = ar.aircraft_code
)

SELECT
  flight_no,
  range
FROM flights_range_rankings
WHERE range_rankings = 1;
```