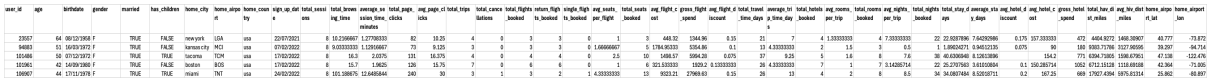
QUERY 1: MAIN TABLE

Zoom In to see the table's head



user_id	age	birthdate	gender	married	has_children	home_city	home_airport	home_country	sign_up_date	total_sessions	total_browsing_time	average_session_time_minutes	total_page_clicks	avg_page_clicks	total_trips	total_cancellations	total_flights_booked	return_flights_booked	single_flights_booked	avg_seats_per_flight	total_seats_booked	avg_flight_cost	gross_flight_spend	avg_flight_discount	total_travel_time_days	average_trip_time_day	total_hotels_booked	avg_rooms_per_trip	total_rooms_booked	avg_nights_per_trip	total_nights_booked	total_stay_days	average_stay_days	avg_hotel_discount	avg_hotel_cost	gross_hotel_spend	total_haversine_miles	avg_flight_miles	home_airport_lat	home_airport_lon
23557	64	08/12/1958	F	TRUE	FALSE	newyork	LGA	usa	22/07/2021	8	10.2166667	1.27708333	82	10.25	4	0	3	3	0	1	3	448.32	1344.96	0.15	21	7	4	1.33333333	4	7.33333333	22	22.8287896	7.64292986	0.175	157.333333	472	4404.9272	1468.30907	40.777	-73.872
94883	51	16/03/1972	F	TRUE	FALSE	kansas city	MCI	usa	07/02/2022	8	9.03333333	1.12916667	73	9.125	3	0	3	3	0	1.66666667	5	1784.95333	5354.86	0.1	13	4.33333333	2	1.5	3	0.5	1	1.89624271	0.94512135	0.075	90	180	9383.71766	3127.90555	39.297	-94.714
101486	58	07/12/1972	F	TRUE	TRUE	tacoma	TCM	usa	17/02/2022	8	16.3	2.0375	131	16.375	4	0	4	4	0	2.5	10	1498.57	5994.28	0.075	37	9.25	5	1.6	8	2	38	40.6306940	8.12613896		114.2	771	4394.71855	1098.67951	47.136	-122.476
101961	42	04/09/1980	F	TRUE	FALSE	boston	BOS	usa	17/02/2022	8	15.7	1.9625	126	15.75	7	0	6	6	0	1	6	321.533333	1929.2	0.13333333	26	4.33333333	7	1	7	3.14285714	22	25.2707563	3.61010084	0.1	150.283714	1052	6712.15128	1118.69188	42.364	-71.005
106967	44	17/11/1978	F	TRUE	TRUE	miami	TNT	usa	24/02/2022	8	101.188675	12.6485844	240	30	3	1	3	2	1	4.33333333	13	9323.21	27969.63	0.15	26	13	4	2	8	8.5	34	34.0807484	8.52018711	0.2	167.25	669	17627.4394	5975.81314	25.862	-80.897

---CTE 1: Cohort Extraction---

```
WITH CohortUsers AS (
    SELECT
        u.user_id,
        u.birthdate,
        u.gender,
        u.married,
        u.has_children,
        COUNT(DISTINCT s.session_id) AS total_sessions,
        u.home_city,
        u.home_country,
        u.home_airport,
        u.home_airport_lat,
        u.home_airport_lon,
        u.sign_up_date
    FROM
        users AS u
    LEFT JOIN
        sessions AS s ON u.user_id = s.user_id
    WHERE
        s.session_start > '2023-01-04 00:00:00'
    GROUP BY
        u.user_id
    HAVING
        COUNT(DISTINCT s.session_id) > 7
),
```

---CTE 2: Latest Session to establish user age---

```
LatestSession AS (
    SELECT
```

```
        user_id,
        MAX(session_end) AS latest_session_end
    FROM
        sessions
    WHERE
        session_start > '2023-01-04 00:00:00'
    GROUP BY
        user_id
),
---CTE 3: Flight-related information---
FlightInfo AS (
    SELECT
        s.user_id,
        COUNT(DISTINCT s.trip_id) AS total_trips,
        SUM(CASE WHEN s.flight_booked THEN 1 ELSE 0 END) AS total_flights_booked,
        SUM(CASE WHEN s.flight_booked AND NOT f.return_flight_booked THEN 1 ELSE 0 END) AS
single_flights_booked,
        SUM(CASE WHEN f.return_flight_booked THEN 1 ELSE 0 END) AS return_flights_booked,
        AVG(f.seats) AS avg_seats_per_flight,
        SUM(f.seats) AS total_seats_booked,
        AVG(f.base_fare_usd) AS avg_flight_cost,
        SUM(f.base_fare_usd) AS gross_flight_spend,
        AVG(s.flight_discount_amount) AS avg_flight_discount,
        SUM(EXTRACT(EPOCH FROM f.return_time - f.departure_time) / (60*60*24)) AS
total_travel_time_days,
        AVG(EXTRACT(EPOCH FROM f.return_time - f.departure_time) / (60*60*24)) AS
average_trip_time_days
    FROM
        sessions s
    LEFT JOIN
        flights f ON s.trip_id = f.trip_id
    GROUP BY
        s.user_id
)
```

```
-- CTE 4: Hotel-related information--
HotelInfo AS (
    SELECT
        s.user_id,
        SUM(CASE WHEN s.hotel_booked THEN 1 ELSE 0 END) AS total_hotels_booked,
        AVG(h.rooms) AS avg_rooms_per_trip,
        SUM(h.rooms) AS total_rooms_booked,
        AVG(h.nights) AS avg_nights_per_trip,
        SUM(h.nights) AS total_nights_booked,
        SUM(EXTRACT(EPOCH FROM h.check_out_time - h.check_in_time) / (60*60*24)) AS
total_stay_days,
        AVG(EXTRACT(EPOCH FROM h.check_out_time - h.check_in_time) / (60*60*24)) AS
average_stay_days,
        AVG(s.hotel_discount_amount) AS avg_hotel_discount,
        AVG(h.hotel_per_room_usd) AS avg_hotel_cost,
        SUM(h.hotel_per_room_usd) AS gross_hotel_spend
    FROM
        sessions s
    LEFT JOIN
        hotels h ON s.trip_id = h.trip_id
    GROUP BY
        s.user_id
),
---CTE 5: Haversine Distance Calculation---
HaversineDistances AS (
    SELECT
        s.user_id,
        f.trip_id,
        f.origin_airport,
        f.destination_airport,
        f.destination_airport_lat AS dest_lat,
        f.destination_airport_lon AS dest_lon,
        u.home_airport_lat AS home_lat,
```

```
        u.home_airport_lon AS home_lon,
        3959 * acos(
            cos(radians(u.home_airport_lat)) * cos(radians(f.destination_airport_lat)) *
cos(radians(f.destination_airport_lon) - radians(u.home_airport_lon)) +
            sin(radians(u.home_airport_lat)) * sin(radians(f.destination_airport_lat))
        ) AS haversine_distance_miles
    FROM
        sessions s
    LEFT JOIN
        flights f ON s.trip_id = f.trip_id
    INNER JOIN
        CohortUsers u ON s.user_id = u.user_id
    WHERE
        u.home_airport_lat IS NOT NULL
        AND u.home_airport_lon IS NOT NULL
        AND f.destination_airport_lat IS NOT NULL
        AND f.destination_airport_lon IS NOT NULL
),
---CTE 6: Average Haversine Distance Calculation---
AvgHaversineDistance AS (
    SELECT
        user_id,
        SUM(haversine_distance_miles) AS total_hav_dist_miles,
        AVG(haversine_distance_miles) AS avg_hiv_dist_miles
    FROM
        HaversineDistances
    GROUP BY
        user_id
)
---FINAL AGGREGATED TABLE---

SELECT
    cu.user_id,
```

```
    --- Age calculation using the latest session end date---
    (DATE_PART('year', ls.latest_session_end) - DATE_PART('year', cu.birthdate) -
    CASE
        WHEN DATE_PART('month', ls.latest_session_end) < DATE_PART('month', cu.birthdate) THEN
1
        WHEN DATE_PART('month', ls.latest_session_end) = DATE_PART('month', cu.birthdate) AND
            DATE_PART('day', ls.latest_session_end) < DATE_PART('day', cu.birthdate) THEN 1
        ELSE 0
    END) AS age,
    ---Other demographic and geographic information---
    cu.birthdate,
    cu.gender,
    cu.married,
    cu.has_children,
    cu.home_city,
    cu.home_airport,
    cu.home_country,
    --- Online activity---
    cu.sign_up_date,
    cu.total_sessions,
    SUM(EXTRACT(EPOCH FROM session_end - session_start) / 60) AS total_browsing_time,
    AVG(EXTRACT(EPOCH FROM session_end - session_start) / 60) AS
average_session_time_minutes,
    SUM(s.page_clicks) AS total_page_clicks,
    AVG(s.page_clicks) AS avg_page_clicks
    ---Trips and cancellations---
    fi.total_trips,
    SUM(CASE WHEN s.cancellation THEN 1 ELSE 0 END) AS total_cancellations,
    ---Flights---
    fi.total_flights_booked,
    fi.return_flights_booked,
    fi.single_flights_booked,
    fi.avg_seats_per_flight,
```

```
    fi.total_seats_booked,
    fi.avg_flight_cost,
    fi.gross_flight_spend,
    fi.avg_flight_discount,
    fi.total_travel_time_days,
    fi.average_trip_time_days,
    ---Hotels---
    hi.total_hotels_booked,
    hi.avg_rooms_per_trip,
    hi.total_rooms_booked,
    hi.avg_nights_per_trip,
    hi.total_nights_booked,
    hi.total_stay_days,
    hi.average_stay_days,
    hi.avg_hotel_discount,
    hi.avg_hotel_cost,
    hi.gross_hotel_spend,
    ---Haversine Distance---
    ad.total_hav_dist_miles,
    ad.avg_hiv_dist_miles,
    ---Geospatial coordinates---
    cu.home_airport_lat,
    cu.home_airport_lon
---End of SELECT---
FROM
    CohortUsers cu
LEFT JOIN
    LatestSession ls ON cu.user_id = ls.user_id
LEFT JOIN
    FlightInfo fi ON cu.user_id = fi.user_id
LEFT JOIN
    HotelInfo hi ON cu.user_id = hi.user_id
LEFT JOIN
```

```
    AvgHaversineDistance ad ON cu.user_id = ad.user_id
LEFT JOIN
    sessions s ON cu.user_id = s.user_id
WHERE
    s.session_start > '2023-01-04 00:00:00'
GROUP BY
    cu.user_id,
    cu.birthdate,
    cu.gender,
    cu.married,
    cu.has_children,
    cu.total_sessions,
    cu.home_city,
    cu.home_country,
    cu.home_airport,
    cu.home_airport_lat,
    cu.home_airport_lon,
    cu.sign_up_date,
    ls.latest_session_end,
    fi.total_trips,
    fi.total_flights_booked,
    fi.return_flights_booked,
    fi.single_flights_booked,
    fi.avg_seats_per_flight,
    fi.total_seats_booked,
    fi.avg_flight_cost,
    fi.gross_flight_spend,
    fi.avg_flight_discount,
    fi.total_travel_time_days,
    fi.average_trip_time_days,
    hi.total_hotels_booked,
    hi.avg_rooms_per_trip,
    hi.total_rooms_booked,
```

```
    hi.avg_nights_per_trip,
    hi.total_nights_booked,
    hi.total_stay_days,
    hi.average_stay_days,
    hi.avg_hotel_discount,
    hi.avg_hotel_cost,
    hi.gross_hotel_spend,
    ad.total_hav_dist_miles,
    ad.avg_hiv_dist_miles
ORDER BY
    cu.user_id;
```

QUERY 2: USER-BASED 'LONG DATA' ON SESSIONS, FLIGHTS, AND HOTEL BOOKINGS

Zoom In to see the table's head

user_id	session_id	trip_id	flight_booked	flight_discount	return_flight_booked	home_airport	origin_airport	destination_airport	destination	home_city	home_country	base_fare_usd	checked_bags	trip_airline	hotel_booked	hotel_discount	hotel_name	nights	rooms	hotel_per_room_usd	cancellation
23557	23557-c24ac	23557-3354(	TRUE	TRUE	TRUE	LGA	LGA	YEG	edmonton	new york	usa	623.25	1	American Airlines	TRUE	TRUE					TRUE
23557	23557-5279c	23557-3354(	TRUE	FALSE	TRUE	LGA	LGA	YEG	edmonton	new york	usa	623.25	1	American Airlines	FALSE	FALSE					FALSE
23557	23557-3d4ef	23557-127a(	TRUE	TRUE	TRUE	LGA	LGA	YKZ	toronto	new york	usa	98.46	1	Ryanair	TRUE	FALSE	Marriott - toronto	2	1	118	FALSE
23557	23557-d1e9(	23557-0c17(	FALSE	FALSE		LGA				new york	usa				TRUE	TRUE	InterContinental - new york	2	1	263	FALSE
23557	23557-1a97(	23557-753d(	FALSE	FALSE		LGA				new york	usa				TRUE	FALSE	Extended Stay - calgary	18	2	91	FALSE

```
WITH UserSessionCounts AS (
    SELECT
        user_id,
        COUNT(DISTINCT session_id) AS session_count
    FROM
        sessions
    WHERE
        session_start > '2023-01-04 00:00:00'
    GROUP BY
        user_id
)
SELECT
    s.user_id,
    s.session_id,
    s.trip_id,
    s.flight_booked,
    s.flight_discount,
```

```
    f.return_flight_booked,
    u.home_airport,
    f.origin_airport,
    f.destination_airport,
    f.destination,
    u.home_city,
    u.home_country,
    f.base_fare_usd,
    f.checked_bags,
    f.trip_airline,
    s.hotel_booked,
    s.hotel_discount,
    h.hotel_name,
    h.nights,
    h.rooms,
    h.hotel_per_room_usd,
    s.cancellation
FROM
    sessions AS s
JOIN
    UserSessionCounts AS usc ON s.user_id = usc.user_id
LEFT JOIN
    flights AS f ON s.trip_id = f.trip_id
LEFT JOIN
                hotels AS h ON s.trip_id = h.trip_id
LEFT JOIN
                users AS u ON s.user_id = u.user_id
WHERE
    usc.session_count > 7
ORDER BY
    s.user_id;
```

QUERY 3: 'LONG DATA' USER SPEND STRUCTURE & CUSTOMER VALUE OVER THE PERIOD (NET BOOKING SPEND)

Zoom In to see the table's head

user_id	base_fare_usd	flight_discount_amount	hotel_per_room_usd	hotel_discount_amount	net_flight_spend	net_hotel_spend	net_booking_spend
23557	98.46	0.15	118		83.691	118	201.691
23557	623.25				1246.5	0	1246.5
23557			91		0	91	91
23557			263	0.25	0	197.25	197.25
23557				0.1	0	0	0

```
WITH UserSessionCounts AS (
    SELECT
        user_id,
        COUNT(DISTINCT session_id) AS session_count
    FROM
        sessions
    WHERE
        session_start > '2023-01-04 00:00:00'
    GROUP BY
        user_id
)
SELECT
    s.user_id,
    f.base_fare_usd,
    s.flight_discount_amount,
    h.hotel_per_room_usd,
    s.hotel_discount_amount,
    SUM(COALESCE(f.base_fare_usd, 0) - (COALESCE(f.base_fare_usd, 0) *
COALESCE(s.flight_discount_amount, 0))) as net_flight_spend,
    SUM(COALESCE(h.hotel_per_room_usd, 0) - (COALESCE(h.hotel_per_room_usd, 0) *
COALESCE(s.hotel_discount_amount, 0))) as net_hotel_spend,
    SUM((COALESCE(f.base_fare_usd, 0) - (COALESCE(f.base_fare_usd, 0) *
COALESCE(s.flight_discount_amount, 0))) + (COALESCE(h.hotel_per_room_usd, 0) -
(COALESCE(h.hotel_per_room_usd, 0) * COALESCE(s.hotel_discount_amount, 0)))) as
net_booking_spend
FROM
    sessions AS s
JOIN
    UserSessionCounts AS usc ON s.user_id = usc.user_id
```

```
LEFT JOIN
    flights AS f ON s.trip_id = f.trip_id
LEFT JOIN
    hotels AS h ON s.trip_id = h.trip_id
LEFT JOIN
    users AS u ON s.user_id = u.user_id
WHERE
    usc.session_count > 7
GROUP BY
    s.user_id,
    f.base_fare_usd,
    s.flight_discount_amount,
    h.hotel_per_room_usd,
    s.hotel_discount_amount
ORDER BY
    user_id ASC;
```