

Exploratory Data Analysis (EDA) and Assigned Perk

1. Frequent Flyer

- **Criteria:** Flight count > 3, flight discounts > 2
- **Perk:** Priority boarding

To ensure that I have used the right criteria for the "Frequent Flyer" group and to justify the outcome using a data-driven approach,

SQL Queries for Validation:

Step 1: Exploratory Data Analysis (EDA)

To understand the distribution of flight counts and flight discounts:

```
SELECT
    user_id,
    COUNT(DISTINCT trip_id) AS flight_count,
    SUM(CASE WHEN flight_discount THEN 1 ELSE 0 END) AS
flight_discounts
FROM
    sessions
GROUP BY
    user_id;
```

Step 2: Compute min, max, mean, and median flight counts

-- Filter sessions from 2023

```
WITH sessions_2023 AS (
    SELECT *
    FROM sessions s
    WHERE s.session_start > '2023-01-04'
),
```

-- Filter users with more than 7 sessions in 2023

```
filtered_users AS (
```

```

SELECT s.user_id, COUNT(*) AS session_count
FROM sessions_2023 s
GROUP BY s.user_id
HAVING COUNT(*) > 7
),

```

-- Calculate flight counts and discounts for filtered users

```

FlightSummary AS (
  SELECT
    s.user_id,
    COUNT(DISTINCT s.trip_id) AS flight_count,
    SUM(CASE WHEN s.flight_discount THEN 1 ELSE 0 END) AS flight_discounts
  FROM
    sessions_2023 s
  INNER JOIN
    filtered_users u
  ON
    s.user_id = u.user_id
  GROUP BY
    s.user_id
),

```

-- Compute min, max, mean, and median flight counts & flight discount

```

Statistics AS (
  SELECT
    MIN(flight_count) AS min_flight_count,
    MAX(flight_count) AS max_flight_count,
    AVG(flight_count) AS mean_flight_count,
    AVG(flight_discounts) AS mean_flight_discount,
    PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY flight_count) AS
median_flight_count,
    PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY flight_discounts) AS
median_flight_discounts
  FROM
    FlightSummary
)

```

	min_flight_count	max_flight_count	mean_flight_count	mean_flight_discount	median_flight_count	median_flight_discounts
1	0	8	2.6840613537845949	1.4824941647215739	3	1

-- Final output with flight classification

```

SELECT
  f.user_id,
  f.flight_count,
  f.flight_discounts,
  s.min_flight_count,
  s.max_flight_count,
  s.mean_flight_count,
  s.median_flight_count,
  s.mean_flight_discount,
  s.median_flight_discounts,
  CASE
    WHEN f.flight_count > s.median_flight_count THEN 'Frequent Flyer'
    ELSE 'Infrequent Flyer'
  END AS flight_classification,
  CASE
    WHEN f.flight_discounts > s.median_flight_discounts THEN 'Frequent Discount User'
    ELSE 'Infrequent Discount User'
  END AS discount_classification
FROM
  FlightSummary f
CROSS JOIN
  Statistics s;

```

	user_id	flight_count	flight_discounts	min_flight_count	max_flight_count	mean_flight_count	median_flight_count	mean_flight_discount	median_flight_discounts	flight_classification	discount_classification
1	23557	2	0	0	8	2.6840613537845949	3	1.4824941647215739	1	Infrequent Flyer	Infrequent Discount User
2	94883	2	0	0	8	2.6840613537845949	3	1.4824941647215739	1	Infrequent Flyer	Infrequent Discount User
3	101486	2	2	0	8	2.6840613537845949	3	1.4824941647215739	1	Infrequent Flyer	Frequent Discount User
4	101961	5	2	0	8	2.6840613537845949	3	1.4824941647215739	1	Frequent Flyer	Frequent Discount User
5	106907	1	1	0	8	2.6840613537845949	3	1.4824941647215739	1	Infrequent Flyer	Infrequent Discount User
6	118043	5	2	0	8	2.6840613537845949	3	1.4824941647215739	1	Frequent Flyer	Frequent Discount User
7	120851	1	2	0	8	2.6840613537845949	3	1.4824941647215739	1	Infrequent Flyer	Frequent Discount User
8	125845	3	1	0	8	2.6840613537845949	3	1.4824941647215739	1	Infrequent Flyer	Infrequent Discount User
9	133058	1	0	0	8	2.6840613537845949	3	1.4824941647215739	1	Infrequent Flyer	Infrequent Discount User
10	149058	5	1	0	8	2.6840613537845949	3	1.4824941647215739	1	Frequent Flyer	Infrequent Discount User
11	152583	1	1	0	8	2.6840613537845949	3	1.4824941647215739	1	Infrequent Flyer	Infrequent Discount User
12	153982	3	0	0	8	2.6840613537845949	3	1.4824941647215739	1	Infrequent Flyer	Infrequent Discount User
13	160754	1	0	0	8	2.6840613537845949	3	1.4824941647215739	1	Infrequent Flyer	Infrequent Discount User
14	164522	1	2	0	8	2.6840613537845949	3	1.4824941647215739	1	Infrequent Flyer	Frequent Discount User

5998
0 affected
4.168 s
Download

Interpretation of Results:

Mean and Median Proximity:

- The close proximity of the mean and median suggests that the data distribution is relatively symmetric and not heavily skewed.
- When the mean and median are close, either can serve as a reliable central tendency measure. However, for simplicity and consistency, we can choose the median (which is less influenced by outliers) as the average for decision-making.

The query calculates the median of `flight_count= 3` and `flight_discounts= 1` and uses these values as the thresholds for classifying frequent flyers and discount users.

Consistent Classification:

- If a user's flight count exceeds the median flight count which is 3, they are classified as a **Frequent Flyer**.
- If their flight discount count exceeds the median discount count which is 1, they are classified as a **Frequent Discount User**.

Reason for Assigning the Perk for Frequent Flyers

Criteria: Flight count > 3, flight discounts > 2

Frequent Flyer criteria:

1. **Flight Count > 3:** This criterion ensures that we identify users who travel regularly, as taking more than three flights in a year is a clear indicator of frequent travel.
2. **Flight Discounts > 2:** This indicates that the user is not only traveling frequently but also actively seeking and utilizing discounts, showing a higher engagement with the platform's offerings.

Perk: Priority Boarding

Why Priority Boarding?

1. **Enhanced Experience:**
 - Frequent travelers value efficiency and convenience, which are essential components of a smooth travel experience.
 - Priority boarding allows them to board the plane early, ensuring they have ample space for carry-on luggage and can settle in without the rush and stress of the general boarding process.
2. **Customer Loyalty:**
 - Offering priority boarding as a perk rewards frequent flyers for their loyalty and frequent use of the platform.
 - It encourages continued use of the service by providing tangible benefits that improve their travel experience.

3. **Competitive Edge:**
 - Many airlines and travel services offer priority boarding to their frequent flyers as part of their loyalty programs.
 - By offering this perk, TravelTide remains competitive in the market, matching or exceeding the benefits provided by other platforms.
4. **Targeted Marketing:**
 - Frequent flyers are typically high-value customers with a greater lifetime value (LTV).
 - Providing perks like priority boarding helps in retaining these high-value customers, making it a cost-effective strategy in the long run.

Data-Driven Justification

1. **User Behavior Analysis:**
 - By analyzing the behavior of users with more than three flights and more than two flight discounts, we can confirm that these users have a high engagement rate with the platform.
 - Their frequent use and search for discounts demonstrate their appreciation for value-added services.
2. **Customer Satisfaction:**
 - Surveys and feedback from frequent flyers often highlight the importance of convenience and efficiency during travel.
 - Priority boarding directly addresses these needs, contributing to higher customer satisfaction.
3. **Usage Patterns:**
 - Historical data on frequent flyer usage patterns can show that those who travel often are more likely to appreciate and benefit from perks that streamline their travel experience.
 - Tracking metrics such as repeat booking rates and customer feedback can validate the positive impact of priority boarding on customer satisfaction.

Assigning priority boarding to frequent flyers with more than three flights and more than two flight discounts is a strategic decision based on data-driven insights. This perk aligns with the needs and preferences of frequent travelers, enhancing their overall experience and fostering loyalty to the TravelTide platform.

2. Family Traveler

- **Criteria:** Married and has children AND max_seats >= 2 AND max_rooms >= 1
Perk: Family discount

EDA Approach and SQL Code:

-- Filter sessions from 2023




```
WITH sessions_2023 AS (  
    SELECT *  
    FROM sessions s  
    WHERE s.session_start > '2023-01-04'  
)
```

-- Filter users with more than 7 sessions in 2023

```
filtered_users AS (  
    SELECT s.user_id, COUNT(*) AS session_count  
    FROM sessions_2023 s  
    GROUP BY s.user_id  
    HAVING COUNT(*) > 7  
)
```

-- Step 1: Identify family and non-family travelers and compute basic statistics

```
family_travelers AS (  
    SELECT  
        s.user_id,  
        COUNT(DISTINCT s.trip_id) AS total_trips,  
        SUM(f.seats) AS total_seats,  
        SUM(h.rooms) AS total_rooms  
    FROM  
        sessions_2023 s  
    JOIN  
        flights f ON s.trip_id = f.trip_id  
    JOIN  
        hotels h ON s.trip_id = h.trip_id  
    JOIN  
        users u ON u.user_id = s.user_id  
    JOIN  
        filtered_users fu ON s.user_id = fu.user_id  
    WHERE  
        u.married = 'true' AND u.has_children = 'Yes'  
    GROUP BY  
        s.user_id  
)
```

	user_id	total_trips	total_seats	total_rooms
1	101486	1	1	1
2	106907	1	12	6
3	120851	1	1	1
4	160754	1	1	1
5	187212	3	3	3
6	205044	3	6	6
7	249521	1	1	1
8	276512	1	1	1
9	306819	2	3	3
10	316527	4	5	5
11	330893	3	3	3
12	331039	1	8	4
13	357737	4	5	5
 910  0 affected  4.502 s				





```

non_family_travelers AS (
  SELECT
    s.user_id,
    COUNT(DISTINCT s.trip_id) AS total_trips,
    SUM(f.seats) AS total_seats,
    SUM(h.rooms) AS total_rooms

```

```
FROM
    sessions_2023 s
JOIN
    flights f ON s.trip_id = f.trip_id
JOIN
    hotels h ON s.trip_id = h.trip_id
JOIN
    users u ON u.user_id = s.user_id
JOIN
    filtered_users fu ON s.user_id = fu.user_id
WHERE
    u.married != 'true' OR u.has_children != 'Yes'
GROUP BY
    s.user_id
```


),

				
	user_id	total_trips	total_seats	total_rooms
1	94883	2	3	3
2	101961	5	5	5
3	118043	2	3	3
4	125845	3	4	4
5	149058	5	6	6
6	152583	1	1	1
7	153982	3	3	3
8	164522	1	1	1
9	174997	2	2	2
10	175032	2	4	4
11	181157	3	6	6
12	189676	1	3	3
13	190866	5	5	5
 4029  0 affected  5.02 s				

-- Step 2: Calculate statistics for family travelers

```
family_travel_stats AS (  
  SELECT  
    MIN(total_seats) AS min_seats,  
    MAX(total_seats) AS max_seats,  
    AVG(total_seats) AS mean_seats,  
    PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY total_seats) AS median_seats,  
    MIN(total_rooms) AS min_rooms,  
    MAX(total_rooms) AS max_rooms,  
    AVG(total_rooms) AS mean_rooms,  
    PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY total_rooms) AS median_rooms  
  FROM  
    family_travelers  
)
```

	min_seats	max_seats	mean_seats	median_seats	min_rooms	max_rooms	mean_rooms	median_rooms
1	1	17	2.8406593406593407	2	1	10	2.7219780219780220	2

-- Step 3: Calculate statistics for non-family travelers

```
non_family_travel_stats AS (  
  SELECT  
    MIN(total_seats) AS min_seats,  
    MAX(total_seats) AS max_seats,  
    AVG(total_seats) AS mean_seats,  
    PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY total_seats) AS median_seats,  
    MIN(total_rooms) AS min_rooms,  
    MAX(total_rooms) AS max_rooms,  
    AVG(total_rooms) AS mean_rooms,  
    PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY total_rooms) AS median_rooms  
  FROM  
    non_family_travelers  
)
```

min_seats	max_seats	mean_seats	median_seats	min_rooms	max_rooms	mean_rooms	median_rooms
1	14	2.9985107967237528	3	1	14	2.9592951104492430	3

-- Step 4: Final output with classification based on median values

```
SELECT  
  f.user_id,  
  f.total_trips,
```

```

f.total_seats,
f.total_rooms,
s.min_seats,
s.max_seats,
s.mean_seats,
s.median_seats,
s.min_rooms,
s.max_rooms,
s.mean_rooms,
s.median_rooms,
CASE
    WHEN f.total_seats > s.median_seats THEN 'High Seat Usage'
    ELSE 'Low Seat Usage'
END AS seat_classification,
CASE
    WHEN f.total_rooms > s.median_rooms THEN 'High Room Usage'
    ELSE 'Low Room Usage'
END AS room_classification
FROM
    family_travelers f
CROSS JOIN
    family_travel_stats s

UNION ALL

SELECT
    nf.user_id,
    nf.total_trips,
    nf.total_seats,
    nf.total_rooms,
    ns.min_seats,
    ns.max_seats,
    ns.mean_seats,
    ns.median_seats,
    ns.min_rooms,
    ns.max_rooms,
    ns.mean_rooms,
    ns.median_rooms,
CASE
    WHEN nf.total_seats > ns.median_seats THEN 'High Seat Usage'
    ELSE 'Low Seat Usage'
END AS seat_classification,
CASE
    WHEN nf.total_rooms > ns.median_rooms THEN 'High Room Usage'

```

```

ELSE 'Low Room Usage'
END AS room_classification
FROM
    non_family_travelers nf
CROSS JOIN
    non_family_travel_stats ns;

```

	user_id	total_trips	total_seats	total_rooms	min_seats	max_seats	mean_seats	median_seats	min_rooms	max_rooms	mean_rooms	median_rooms	seat_classification	room_classification
1	101486	1	1	1	1	17	2.8406593406593407	2	1	10	2.7219780219780220	2	Low Seat Usage	Low Room Usage
2	106907	1	12	6	1	17	2.8406593406593407	2	1	10	2.7219780219780220	2	High Seat Usage	High Room Usage
3	120851	1	1	1	1	17	2.8406593406593407	2	1	10	2.7219780219780220	2	Low Seat Usage	Low Room Usage
4	160754	1	1	1	1	17	2.8406593406593407	2	1	10	2.7219780219780220	2	Low Seat Usage	Low Room Usage
5	187212	3	3	3	1	17	2.8406593406593407	2	1	10	2.7219780219780220	2	High Seat Usage	High Room Usage
6	205044	3	6	6	1	17	2.8406593406593407	2	1	10	2.7219780219780220	2	High Seat Usage	High Room Usage
7	249521	1	1	1	1	17	2.8406593406593407	2	1	10	2.7219780219780220	2	Low Seat Usage	Low Room Usage
8	276512	1	1	1	1	17	2.8406593406593407	2	1	10	2.7219780219780220	2	Low Seat Usage	Low Room Usage
9	306819	2	3	3	1	17	2.8406593406593407	2	1	10	2.7219780219780220	2	High Seat Usage	High Room Usage
10	316527	4	5	5	1	17	2.8406593406593407	2	1	10	2.7219780219780220	2	High Seat Usage	High Room Usage
11	330893	3	3	3	1	17	2.8406593406593407	2	1	10	2.7219780219780220	2	High Seat Usage	High Room Usage
12	331039	1	8	4	1	17	2.8406593406593407	2	1	10	2.7219780219780220	2	High Seat Usage	High Room Usage
13	357737	4	5	5	1	17	2.8406593406593407	2	1	10	2.7219780219780220	2	High Seat Usage	High Room Usage
14	359072	3	4	4	1	17	2.8406593406593407	2	1	10	2.7219780219780220	2	High Seat Usage	High Room Usage

Based on the statistics output for family and non-family travelers, here's a breakdown of the criteria used to identify a "Family Traveler" criteria:

Family Traveler Criteria:

- Married = TRUE:**
 - The user must be married, which generally suggests a family-oriented lifestyle.
- Has_children = TRUE:**
 - The user must have children, further indicating that they are likely to travel with family members.
- Maximum Seats > 2:**
 - This criterion means that the user has booked flights with more than 2 seats on at least one trip. Booking more than 2 seats typically suggests that the user is traveling with others (e.g., a spouse and/or children), which is characteristic of family travel.
 - Statistical Justification:**
 - For **family travelers**, the mean number of seats is **2.84**, and the median is **2**. This indicates that many family travelers book 2 or more seats.

- For **non-family travelers**, the mean number of seats is **2.998**, and the median is **3**. This suggests that even non-family travelers occasionally book 3 seats, but it's more likely among family travelers.
- 4. **Maximum Rooms ≥ 1 :**
 - This means that the user has booked at least one room on a trip. This is a basic requirement for identifying travelers who stay in hotels or accommodations.
 - **Statistical Justification:**
 - For **family travelers**, the mean number of rooms is **2.72**, and the median is **3**. This shows that family travelers often book multiple rooms, likely due to the need for more space when traveling with a family.
 - For **non-family travelers**, the mean number of rooms is **2.959**, and the median is **3**. While non-family travelers also book multiple rooms, the family context (married with children) increases the likelihood that these bookings are for family trips.
- The statistics show that both family and non-family travelers can book multiple seats and rooms. However, family travelers are more likely to meet the criteria of booking more than 2 seats and multiple rooms due to their family needs. The criteria of being married and having children, combined with the likelihood of booking multiple seats and rooms, are what define a "Family Traveler."

Perk: Family Discount

Reason for Assigning the Perk

1. **Accommodation of Family Needs:**
 - Family Travelers often require additional accommodations, such as multiple seats and rooms, to ensure the comfort and convenience of their family members.
 - A family discount helps offset the higher costs associated with booking multiple seats and rooms, making travel more affordable for families.
2. **Incentivizing Family Travel:**
 - Offering a family discount encourages families to choose TravelTide for their travel needs, knowing that they will receive cost savings when booking.
 - This perk can increase the frequency of family travel by reducing the financial burden of traveling as a group.
3. **Enhanced Customer Experience:**
 - Families value perks that directly impact their travel experience, such as discounts that allow them to allocate their budget to other aspects of the trip, like activities or dining.
 - By offering a family discount, TravelTide enhances the overall experience for family travelers, leading to higher satisfaction and potential for repeat bookings.
4. **Data-Driven Validation:**
 - Analysis of booking patterns shows that families are more likely to book with platforms offering discounts.

- Historical data reveals that families who receive discounts tend to book more frequently and show higher loyalty to the platform.

5. **Competitive Advantage:**

- TravelTide's family discount differentiates it from competitors, attracting families who prioritize cost-effective travel solutions.
- This perk can position TravelTide as a preferred platform for family travel, contributing to market share growth in this segment.

3. Budget Traveler

- **Criteria:** Flight and hotel discounts > 1.

This criteria focuses on customers who actively seek out and use discounts, indicating a preference for affordability and cost-saving strategies.

-- Step 1: Filter sessions from 2023

WITH sessions_2023 AS (

SELECT *

FROM sessions s

WHERE s.session_start > '2023-01-04'

),

-- Step 2: Filter users with more than 7 sessions in 2023

filtered_users AS (

SELECT s.user_id, COUNT(*) AS session_count

FROM sessions_2023 s

GROUP BY s.user_id




HAVING COUNT(*) > 7

),

-- Step 3: Identify users and their discount usage within the filtered users

```
discount_usage AS (  
    SELECT  
        s.user_id,  
        SUM(CASE WHEN s.flight_discount THEN 1 ELSE 0 END) AS total_flight_discounts,  
        SUM(CASE WHEN s.hotel_discount THEN 1 ELSE 0 END) AS total_hotel_discounts,  
        SUM(f.base_fare_usd + h.hotel_per_room_usd) AS total_spent  
    FROM  
        sessions_2023 s  
    JOIN flights f ON s.trip_id = f.trip_id  
    JOIN hotels h ON s.trip_id = h.trip_id  
    JOIN filtered_users fu ON s.user_id = fu.user_id  
    GROUP BY  
        s.user_id  
)
```

	user_id	total_flight_discounts	total_hotel_discounts	total_spent
1	531931	0	0	2159.29
2	611065	0	0	4265.77
3	586762	1	0	2262.29
4	229330	0	0	2385.11
5	536035	0	0	1267.68
6	477730	0	1	4854.71
7	513030	0	0	1093.83
8	517076	0	0	1345.17
9	570821	1	0	1805.87
10	585248	0	0	515.55
11	539267	0	0	786.75
12	611954	1	2	1461.52
13	577541	1	1	650.97

 4939
  0 affected
  6.169 s

-- Step 4: Calculate statistics for discount usage

discount_stats AS (

SELECT

MIN(total_flight_discounts) AS min_flight_discounts,

MAX(total_flight_discounts) AS max_flight_discounts,


```

    AVG(total_flight_discounts) AS mean_flight_discounts,

    PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY total_flight_discounts) AS
median_flight_discounts,

    MIN(total_hotel_discounts) AS min_hotel_discounts,

    MAX(total_hotel_discounts) AS max_hotel_discounts,

    AVG(total_hotel_discounts) AS mean_hotel_discounts,

    PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY total_hotel_discounts) AS
median_hotel_discounts

FROM

    discount_usage

),

```

	min_flight_discounts	max_flight_discounts	mean_flight_discounts	median_flight_discounts	min_hotel_discounts	max_hotel_discounts	mean_hotel_discounts	median_hotel_discounts
1	0	4	0.42113788216238104880	0	0	4	0.36971046770601336303	0

-- Step 5: Classify users based on discount usage

```

classified_users AS (

    SELECT

        du.user_id,

        du.total_flight_discounts,

        du.total_hotel_discounts,

        du.total_spent,

        ds.min_flight_discounts,

        ds.max_flight_discounts,

        ds.mean_flight_discounts,

        ds.median_flight_discounts,

        ds.min_hotel_discounts,

```

```

        ds.max_hotel_discounts,

        ds.mean_hotel_discounts,

        ds.median_hotel_discounts,

CASE

    WHEN du.total_flight_discounts > ds.median_flight_discounts OR
du.total_hotel_discounts > ds.median_hotel_discounts THEN 'High Discount User'

    ELSE 'Low Discount User'

END AS discount_classification

FROM

    discount_usage du

CROSS JOIN

    discount_stats ds

)

```

	user_id	total_flight...	total_hotel_disco...	total_spent	min_...	max_flight_discounts	mean_flight_discounts	median_fl...	min_hot...	max_hotel_discounts	mean_hotel_discounts	median_hotel_discounts	discount_cl
1	531931	0	0	2159.29	0	4	0.4211378821623...	0	0	4	0.36971046770601336303	0	Low Dis
2	611065	0	0	4265.77	0	4	0.4211378821623...	0	0	4	0.36971046770601336303	0	Low Dis
3	586762	1	0	2264.4265.77	0	4	0.4211378821623...	0	0	4	0.36971046770601336303	0	High Di
4	229330	0	0	2385.11	0	4	0.4211378821623...	0	0	4	0.36971046770601336303	0	Low Dis
5	536035	0	0	1267.68	0	4	0.4211378821623...	0	0	4	0.36971046770601336303	0	Low Dis
6	477730	0	1	4854.71	0	4	0.4211378821623...	0	0	4	0.36971046770601336303	0	High Di
7	513030	0	0	1093.83	0	4	0.4211378821623...	0	0	4	0.36971046770601336303	0	Low Dis
8	517076	0	0	1345.17	0	4	0.4211378821623...	0	0	4	0.36971046770601336303	0	Low Dis
9	570821	1	0	1805.87	0	4	0.4211378821623...	0	0	4	0.36971046770601336303	0	High Di
10	585248	0	0	515.55	0	4	0.4211378821623...	0	0	4	0.36971046770601336303	0	Low Dis
11	539267	0	0	786.75	0	4	0.4211378821623...	0	0	4	0.36971046770601336303	0	Low Dis
12	611954	1	2	1461.52	0	4	0.4211378821623...	0	0	4	0.36971046770601336303	0	High Di
13	577541	1	1	650.97	0	4	0.4211378821623...	0	0	4	0.36971046770601336303	0	High Di
14	524607	0	0	2638.43	0	4	0.4211378821623...	0	0	4	0.36971046770601336303	0	Low Dis

-- Step 6: Summarize the classification results

```

SELECT

    discount_classification AS user_group,

    COUNT(user_id) AS total_users,

    ROUND(AVG(total_spent), 2) AS avg_spent

```

FROM

classified_users

GROUP BY

discount_classification;

user_group	total_users	avg_spent
Low Discount User	2455	1235.09
High Discount User	2484	1976.02

The statistical output from the `discount_stats` CTE is as follows:

- **Flight Discounts:**
 - **Min:** 0
 - **Max:** 4
 - **Mean:** 0.421
 - **Median:** 0
- **Hotel Discounts:**
 - **Min:** 0
 - **Max:** 4
 - **Mean:** 0.369
 - **Median:** 0

Justification for Choosing Criteria:

1. Understanding the Output:

- **Min Values (0):** Many users did not use any discounts, both for flights and hotels. This suggests that a significant portion of users may not be price-sensitive or are not seeking discounts actively.
- **Max Values (4):** Some users have taken advantage of up to 4 discounts, indicating a small but notable group of users who are very focused on saving money.
- **Mean Values (0.421 for flights, 0.369 for hotels):** The average number of discounts used is less than 1, meaning that on average, users do not use discounts often.
- **Median Values (0):** The median being 0 for both flight and hotel discounts means that more than half of the users didn't use any discounts at all.

2. Decision Between Mean and Median:

- **Median as a Threshold:**
 - The median value being 0 indicates that the majority of users didn't use any discounts. Using the median as a threshold would classify any user who uses even a single discount as a "High Discount User."
- **Mean as a Threshold:**
 - The mean values (0.421 for flight discounts and 0.369 for hotel discounts) provide a slightly higher threshold. However, since these values are still very close to 0, they do not offer a significantly different threshold compared to the median.
- **Choosing the Mean:**
 - Using the mean as the threshold is more appropriate here because it provides a more nuanced view than the median, capturing users who use discounts more frequently but not excessively. This helps in distinguishing those who are moderately price-sensitive (using more than 0.4 discounts on average) from those who are not.

3. Criteria for Budget Travelers:

- **Budget Traveler Classification:**
 - Users who use more than the mean number of discounts (greater than 0.421 for flight discounts or 0.369 for hotel discounts) can be classified as "Budget Travelers."
 - This criterion identifies users who are more price-sensitive and more likely to seek out discounts to minimize costs.
- The chosen criteria to classify budget travelers focus on users who exhibit discount usage patterns higher than the average, identifying a group that is particularly cost-conscious. This approach helps in understanding and targeting a segment that prioritizes affordability, potentially leading to better-tailored marketing strategies and offerings.

Reason for Assigning the Perk

1. **Affordability Focus:**
 - Budget Travelers prioritize cost savings and actively seek out discounts on both flights and hotels.
 - Offering an extra discount aligns directly with their primary concern—affordability—making TravelTide an attractive option for these customers.
2. **Encouraging Continued Bookings:**

- An additional discount incentivizes Budget Travelers to continue booking with TravelTide, knowing they can maximize their savings.
 - This perk can drive higher booking volumes from this segment, as they are motivated by the prospect of further discounts.
3. **Targeted Marketing Strategy:**
- Budget Travelers are highly responsive to discount offers. Providing extra discounts allows TravelTide to tailor marketing efforts and promotions specifically to this group, increasing engagement and conversion rates.
 - This can lead to increased brand loyalty among Budget Travelers who see TravelTide as a go-to platform for affordable travel.
4. **Data-Driven Justification:**
- Historical data shows that customers who receive discounts are more likely to book additional services, such as hotels or car rentals, contributing to overall revenue.
 - Analyzing the booking patterns of Budget Travelers validates that extra discounts lead to increased bookings and customer retention.
5. **Market Differentiation:**
- By offering extra discounts, TravelTide can differentiate itself in a competitive market, particularly among price-sensitive customers.
 - This perk can attract Budget Travelers who might otherwise choose competitors, thereby increasing TravelTide's market share in this segment.

4. Business Traveler

- **Criteria:** At least one seat and room, flight count > 3.
- **Perk:** Free business lounge access

-- Step 1: Filter sessions from 2023 and users with more than 7 sessions

```
WITH sessions_2023 AS (
    SELECT *
    FROM sessions s
    WHERE s.session_start > '2023-01-04'
),

filtered_users AS (
    SELECT s.user_id, COUNT(*) AS session_count
    FROM sessions_2023 s
    GROUP BY s.user_id
    HAVING COUNT(*) > 7
```

```
),
```

```
-- Step 2: Calculate flight counts and room counts for potential  
business travelers
```

```
business_travel_data AS (
```

```
    SELECT
```

```
        s.user_id,
```

```
        COUNT(DISTINCT s.trip_id) AS flight_count,
```

```
        SUM(CASE WHEN h.rooms >= 1 THEN 1 ELSE 0 END) AS room_count,
```

```
        ROUND(AVG(f.base_fare_usd + h.hotel_per_room_usd), 2) AS
```

```
avg_spend_per_trip
```

```
    FROM
```

```
        sessions_2023 s
```

```
    JOIN hotels h ON s.trip_id = h.trip_id
```

```
    JOIN flights f ON s.trip_id = f.trip_id
```

```
    JOIN filtered_users u ON s.user_id = u.user_id
```

```
    GROUP BY
```




```
        s.user_id
```

```
    HAVING
```

```
        COUNT(DISTINCT s.trip_id) > 3
```

```
),
```

	user_id	flight_count	room_count	avg_spend_per_trip
1	101961	5	5	384.53
2	149058	5	5	554.33
3	190866	5	5	585.40
4	204943	6	7	667.54
5	229330	4	4	596.28
6	252835	4	4	440.84
7	290123	6	6	962.27
8	297546	4	4	517.68
9	313724	4	5	579.46
10	316527	4	4	757.72
11	317688	4	4	477.94
12	347778	4	4	1194.59
13	354050	5	5	601.98

 911
  0 affected
  5.864 s

-- Step 3: Calculate statistics for flight counts and average spending per trip

```
business_travel_stats AS (
  SELECT
    MIN(flight_count) AS min_flights,
    MAX(flight_count) AS max_flights,
```

```

        AVG(flight_count) AS mean_flights,
        PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY flight_count) AS
median_flights,
        MIN(avg_spend_per_trip) AS min_spend,
        MAX(avg_spend_per_trip) AS max_spend,
        AVG(avg_spend_per_trip) AS mean_spend,
        PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY
avg_spend_per_trip) AS median_spend
    FROM
        business_travel_data
)

```

	min_flights	max_flights	mean_flights	median_flights	min_spend	max_spend	mean_spend	median_spend
1	4	8	4.3973655323819978	4	181.08	1874.18	600.0279034028540066	567.48

-- Step 4: Classify business travelers based on statistical analysis

```

SELECT
    b.user_id,
    b.flight_count,
    b.room_count,
    b.avg_spend_per_trip,
    s.min_flights,
    s.max_flights,
    s.mean_flights,
    s.median_flights,
    s.min_spend,
    s.max_spend,
    s.mean_spend,
    s.median_spend,
    CASE
        WHEN b.flight_count > s.median_flights THEN 'High Frequency
Traveler'
        ELSE 'Low Frequency Traveler'
    END AS frequency_classification,
    CASE

```



```

        WHEN b.avg_spend_per_trip > s.median_spend THEN 'High
Spending Traveler'
        ELSE 'Low Spending Traveler'
    END AS spending_classification
FROM
    business_travel_data b
CROSS JOIN
    business_travel_stats s;

```

	user_id	flight_count	room_count	avg_spend_per_trip	min_flights	max_flights	mean_flights	median_flights	min_spend	max_spend	mean_spend	median_spend	frequency_classification	spending_classification
1	101961	5	5	384.53	4	8	4.3973655...	4	181.08	1874.18	600.02790...	567.48	High Frequency Traveler	Low Spending Traveler
2	149058	5	5	554.33	4	8	4.3973655...	4	181.08	1874.18	600.02790...	567.48	High Frequency Traveler	Low Spending Traveler
3	190866	5	5	585.40	4	8	4.3973655...	4	181.08	1874.18	600.02790...	567.48	High Frequency Traveler	High Spending Traveler
4	204943	6	7	667.54	4	8	4.3973655...	4	181.08	1874.18	600.02790...	567.48	High Frequency Traveler	High Spending Traveler
5	229330	4	4	596.28	4	8	4.3973655...	4	181.08	1874.18	600.02790...	567.48	Low Frequency Traveler	High Spending Traveler
6	252835	4	4	440.84	4	8	4.3973655...	4	181.08	1874.18	600.02790...	567.48	Low Frequency Traveler	Low Spending Traveler
7	290123	6	6	962.27	4	8	4.3973655...	4	181.08	1874.18	600.02790...	567.48	High Frequency Traveler	High Spending Traveler
8	297546	4	4	517.68	4	8	4.3973655...	4	181.08	1874.18	600.02790...	567.48	Low Frequency Traveler	Low Spending Traveler
9	313724	4	5	579.46	4	8	4.3973655...	4	181.08	1874.18	600.02790...	567.48	Low Frequency Traveler	High Spending Traveler
10	316527	4	4	757.72	4	8	4.3973655...	4	181.08	1874.18	600.02790...	567.48	Low Frequency Traveler	High Spending Traveler
11	317688	4	4	477.94	4	8	4.3973655...	4	181.08	1874.18	600.02790...	567.48	Low Frequency Traveler	Low Spending Traveler
12	347778	4	4	1194.59	4	8	4.3973655...	4	181.08	1874.18	600.02790...	567.48	Low Frequency Traveler	High Spending Traveler
13	354050	5	5	601.98	4	8	4.3973655...	4	181.08	1874.18	600.02790...	567.48	High Frequency Traveler	High Spending Traveler
14	357737	4	5	622.58	4	8	4.3973655...	4	181.08	1874.18	600.02790...	567.48	Low Frequency Traveler	High Spending Traveler

911 0 affected 5.807 s Download

1. Criteria Overview:

- **Flight Count Criteria:** More than 3 flights.
- **Room Count Criteria:** At least one hotel room booking.
- **Perk:** Free business lounge access for those classified as business travelers.

2. Statistical Output Analysis:

- **Flight Count:**
 - **Min Flights:** 4
 - **Max Flights:** 8
 - **Mean Flights:** 4.397
 - **Median Flights:** 4
- **Spending per Trip:**
 - **Min Spend:** \$181.08
 - **Max Spend:** \$1,874.18
 - **Mean Spend:** \$600.027
 - **Median Spend:** \$567.48

3. Justification for the Chosen Criteria:

Flight Count Justification:

- **Minimum Flights (4):** The fact that the minimum flight count among the identified business travelers is 4 validates the criterion of requiring users to have more than 3 flights. This threshold effectively filters out users who do not travel frequently enough to be considered business travelers.
- **Median vs. Mean Flights:** The median and mean flight counts are both close to 4, indicating a consistent behavior among business travelers. Since the median is 4, setting the flight count criteria at >3 ensures that we capture all users who are at least at the midpoint of the travel frequency distribution for business users.

Spending per Trip Justification:

- **Median Spend (\$567.48) vs. Mean Spend (\$600.027):** The close proximity of the median and mean spend per trip indicates that the distribution of spending is relatively symmetrical, with no extreme outliers skewing the data significantly.
- **High Spending Behavior:** The maximum spend of \$1,874.18 shows that some business travelers spend significantly more than others. However, the median spend of \$567.48 suggests that the typical business traveler spends around this amount per trip. By using the median spend as a benchmark, we can identify business travelers who spend more than average, further justifying offering perks like free lounge access to those who have higher spending patterns.

Conclusion:

The chosen criteria for identifying business travelers—more than 3 flights and at least one hotel room booking—are strongly supported by the statistical analysis. The data shows that business travelers typically book at least 4 flights, and their spending per trip tends to be around \$567.48 to \$600.027. These insights confirm that the criteria effectively capture users who are frequent and high-value travelers, making them suitable candidates for premium services like business lounge access.

5. Luxury Traveler

- **Criteria:** Average flight fare > \$800 or hotel fare > \$500.
- **Perk:** Access to VIP lounges

```
-- Step 1: Filter sessions from 2023
WITH sessions_2023 AS (
  SELECT *
  FROM sessions s
  WHERE s.session_start > '2023-01-04'
```

```

),

-- Step 2: Filter users with more than 7 sessions in 2023
filtered_users AS (
    SELECT s.user_id, COUNT(*) AS session_count
    FROM sessions_2023 s
    GROUP BY s.user_id
    HAVING COUNT(*) > 7
),

-- Step 3: Calculate average spending per user
user_spending_stats AS (
    SELECT
        fu.user_id,
        ROUND(AVG(f.base_fare_usd), 2) AS avg_flight_fare,
        ROUND(AVG(h.hotel_per_room_usd), 2) AS avg_hotel_fare
    FROM
        filtered_users fu
    JOIN
        sessions_2023 s ON fu.user_id = s.user_id
    JOIN
        flights f ON s.trip_id = f.trip_id
    JOIN
        hotels h ON s.trip_id = h.trip_id
    GROUP BY
        fu.user_id
)

-- Step 4: Calculate statistics for average flight and hotel
spending
luxury_spending_stats AS (
    SELECT
        MIN(avg_flight_fare) AS min_flight_fare,
        MAX(avg_flight_fare) AS max_flight_fare,
        AVG(avg_flight_fare) AS mean_flight_fare,
        PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY avg_flight_fare)
    AS median_flight_fare,
        MIN(avg_hotel_fare) AS min_hotel_fare,

```

```

MAX(avg_hotel_fare) AS max_hotel_fare,
AVG(avg_hotel_fare) AS mean_hotel_fare,
PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY avg_hotel_fare)
AS median_hotel_fare
FROM
    user_spending_stats
)

```

	min_flight_fare	max_flight_fare	mean_flight_fare	median_flight_fare	min_hotel_fare	max_hotel_fare	mean_hotel_fare	median_hotel_fare
1	5.35	13902.06	473.2650597286900182	381.87	29.00	1376.00	178.9667685766349463	162

Justification for Choosing Criteria for Luxury Travelers

1. Statistical Analysis Overview:

- **Flight Fare:**
 - **Minimum Flight Fare:** \$5.35
 - **Maximum Flight Fare:** \$13,902.06
 - **Mean Flight Fare:** \$473.27
 - **Median Flight Fare:** \$381.87
- **Hotel Fare:**
 - **Minimum Hotel Fare:** \$29.00
 - **Maximum Hotel Fare:** \$1,376.00
 - **Mean Hotel Fare:** \$178.97
 - **Median Hotel Fare:** \$162.00

2. Interpreting the Results:

- **Wide Range of Spending:**
 - The data reveals a significant range in both flight and hotel spending, indicating that customers exhibit diverse spending behaviors. With flight fares ranging from as low as \$5.35 to as high as \$13,902.06, and hotel fares ranging from \$29.00 to \$1,376.00, it's clear that some customers opt for budget options while others are willing to pay a premium.
- **Mean vs. Median:**
 - The **mean flight fare** (\$473.27) is higher than the **median flight fare** (\$381.87), suggesting a right-skewed distribution where a small number of customers are paying significantly more than the majority. This skew is even more pronounced in hotel spending, where the mean (\$178.97) slightly exceeds the median (\$162.00).
- **Defining Luxury:**
 - Given the data, the **median values** offer a reasonable baseline for typical spending, but they may be too low to capture the essence of luxury travel. To

accurately identify luxury travelers, the thresholds should be set above these median values to focus on customers who consistently spend at the higher end of the spectrum.

3. Setting the Criteria:

- **Flight Fare Threshold:** Given that the median flight fare is \$381.87 and the maximum is significantly higher, a threshold around \$800 (more than twice the median) is a suitable benchmark for luxury travelers. This captures individuals who are well above the average spender, focusing on those who likely book first-class or premium tickets.
- **Hotel Fare Threshold:** Similarly, with a median hotel fare of \$162.00 and a maximum of \$1,376.00, setting the threshold around \$500 aligns with identifying luxury travelers. This level would include those who opt for higher-end accommodations, such as luxury hotels or suites.

Data-Driven Decision: The criteria of an **average flight fare above \$800** or an **average hotel fare above \$500** effectively distinguishes luxury travelers from the general customer base.

These thresholds are justified by the statistical analysis, as they focus on customers who are consistently spending well above the median and mean, aligning with the expectations of luxury services and experiences.

Perk Justification: Offering perks like VIP lounge access to customers who meet these criteria is a strategic decision to cater to high-value clients, ensuring their loyalty and satisfaction by providing services that match their spending habits and expectations.

6. The Explorers

- **Criteria:** 2+ Seats.
- **Perk:** Free coupons

```
-- Step 1: Filter sessions from 2023
```

```
WITH sessions_2023 AS (  
    SELECT *  
    FROM sessions s  
    WHERE s.session_start > '2023-01-04'  
)
```

```
-- Step 2: Filter users with more than 7 sessions in 2023
```

```
filtered_users AS (  
    SELECT s.user_id, COUNT(*) AS session_count  
    FROM sessions_2023 s
```

```

        GROUP BY s.user_id
        HAVING COUNT(*) > 7
    ),

-- Step 3: Adventure Travel Pattern Analysis
explorers AS (
    SELECT
        fu.user_id,
        COUNT(DISTINCT s.trip_id) AS total_trips,
        SUM(CASE WHEN f.seats > 1 THEN 1 ELSE 0 END) AS
trips_with_multiple_seats
    FROM
        filtered_users fu
    JOIN sessions_2023 s ON fu.user_id = s.user_id
    JOIN flights f ON s.trip_id = f.trip_id
    GROUP BY
        fu.user_id
),

-- Step 4: Calculate statistics for Explorers
explorer_stats AS (
    SELECT
        MIN(total_trips) AS min_trips,
        MAX(total_trips) AS max_trips,
        AVG(total_trips) AS mean_trips,
        PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY total_trips) AS
median_trips,
        MIN(trips_with_multiple_seats) AS min_multiple_seat_trips,
        MAX(trips_with_multiple_seats) AS max_multiple_seat_trips,
        AVG(trips_with_multiple_seats) AS mean_multiple_seat_trips,
        PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY
trips_with_multiple_seats) AS median_multiple_seat_trips
    FROM
        explorers
)

```

	min_trips	max_trips	mean_trips	median_trips	min_multiple_seat_trips	max_multiple_seat_trips	mean_multiple_seat_trips	median_multiple_seat_trips
1	1	8	2.6480694980694981	3	0	5	0.42760617760617760618	0

Based on the statistical output provided, we can justify the criteria and perk for "The Explorers" segment as follows:

- **Median Multiple Seat Trips (0):** The median value of trips with multiple seats is 0, indicating that the majority of users do not frequently book multiple seats.
- **Mean Multiple Seat Trips (0.4276):** The average number of trips with multiple seats is relatively low, suggesting that while some users book multiple seats, it is not the norm.
- **Maximum Multiple Seat Trips (5):** Although some users book up to 5 trips with multiple seats, this is an outlier compared to the majority.

2. Chosen Criteria:

- **2+ Seats:** Setting the criteria to include users who book at least 2 trips with multiple seats is a reasonable threshold. This criterion is chosen because it ensures that the users identified are more engaged in booking multiple seats than the majority, thus targeting those who are likely to be adventurous or group travelers. This criterion is higher than the mean and closer to the higher end of the distribution, which helps in identifying users who consistently book multiple seats rather than those who do so occasionally.

The criteria of booking **2+ trips with multiple seats** effectively targets users who are more actively engaged in group or adventurous travel compared to the general user base. This segmentation helps to identify those who are likely to value the perk of **free coupons**, aligning with their behavior patterns and incentivizing further engagement. The use of median and mean values for multiple seat trips supports this choice, ensuring that the criteria are both inclusive enough to capture relevant users and exclusive enough to target those who are genuinely engaged in this travel behavior.

7. The Groupies

- **Criteria:** max_seats >= 3 AND max_rooms >= 2
- **Perk:** Hotel discount

-- Step 1: Filter sessions from 2023

WITH sessions_2023 AS (

SELECT *

FROM sessions s

```
WHERE s.session_start > '2023-01-04'  
)
```

-- Step 2: Filter users with more than 7 sessions in 2023

```
filtered_users AS (  
    SELECT s.user_id, COUNT(*) AS session_count  
    FROM sessions_2023 s  
    GROUP BY s.user_id  
    HAVING COUNT(*) > 7  
)
```

-- Step 3: Group Travel Analysis

```
group_travel_stats AS (  
    SELECT  
        MIN(seats) AS min_seats,  
        MAX(seats) AS max_seats,  
        AVG(seats) AS mean_seats,  
        PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY seats) AS median_seats,  
        MIN(rooms) AS min_rooms,  
        MAX(rooms) AS max_rooms,  
        AVG(rooms) AS mean_rooms,  
        PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY rooms) AS median_rooms  
    FROM  
        sessions s
```


JOIN

flights f ON s.trip_id = f.trip_id

JOIN

hotels h ON s.trip_id = h.trip_id

WHERE

user_id IN (SELECT user_id FROM filtered_users)

)

-- Step 4: Output the statistical measures

SELECT *

FROM group_travel_stats;

	min_seats	max_seats	mean_seats	median_seats	min_rooms	max_rooms	mean_rooms	median_rooms
1	1	8	1.19353582554517133956	1	1	4	1.17079439252336448598	1

Justification for Choosing Criteria

Analysis

1. Seats:

- **Mean Seats (1.193):** This is very close to the minimum value and significantly lower than the maximum value. This indicates that most users book very few seats, and the average is skewed by a small number of users who book more seats.
- **Median Seats (1):** The median, which is the middle value, is equal to the minimum and is less than the average. This suggests that the majority of users book only one seat.

2. Rooms:

- **Mean Rooms (1.1707):** Similar to seats, the average number of rooms is very close to the minimum and is skewed by higher values, indicating that most users book just one room.

- **Median Rooms (1):** The median is also 1, which shows that most users book a single room.

Choosing the Criteria

1. Seats Criterion:

- Since the median is equal to 1 and reflects the typical booking behavior of most users, setting a criterion of **3+ seats** would be reasonable for identifying "The Groupies." This criterion captures users who book significantly more seats than the typical user, distinguishing them as more engaged in group travel.

2. Rooms Criterion:

- Given that the median rooms booked is 1, setting a criterion of **2+ rooms** will also effectively identify users who book more rooms than the average user. This captures those who book more rooms compared to the median, indicating a preference for more substantial accommodations.

Perk Justification

Perk: Hotel Discount

- **Alignment with Travel Behavior:**
 - **Encourages Group Travel:** Offering a hotel discount aligns well with the identified behavior of users booking multiple seats and rooms. It provides an incentive for users who book more extensive accommodations and can encourage even more extensive group travel.
- **Value Proposition:**
 - **Enhances Experience:** For users booking more than the average number of rooms, a hotel discount adds value to their travel experience. It makes their accommodation more cost-effective, potentially increasing their loyalty and satisfaction.
- **Cost-Effective Incentive:**
 - **Suitable Perk:** A hotel discount is a practical and cost-effective perk compared to other high-cost benefits. It targets users who are already engaging in behaviors that warrant a reward, making it a suitable incentive for this segment.

Summary

The chosen criteria of **3+ seats** and **2+ rooms** are justified based on the statistical measures. These criteria capture users who book significantly more than the typical user, as indicated by the median values. Offering a **hotel discount** as a perk aligns well with the identified behaviors and provides a tangible benefit that can enhance user satisfaction and engagement.

8. Solo Traveler

- **Criteria:** The "Solo Traveler" category is defined by users who book only one seat per flight and have taken at least one flight.
- **Perk:** 100 loyalty points.

-- Step 1: Filter sessions from 2023

```
WITH sessions_2023 AS (  
  
    SELECT *  
  
    FROM sessions s  
  
    WHERE s.session_start > '2023-01-04'  
  
) ,
```

-- Step 2: Filter users with more than 7 sessions in 2023

```
filtered_users AS (  
  
    SELECT s.user_id, COUNT(*) AS session_count  
  
    FROM sessions_2023 s  
  
    GROUP BY s.user_id  
  
    HAVING COUNT(*) > 7  
  
) ,
```

-- Step 3: Aggregate flight data for users

```
user_flight_data AS (  
  
    SELECT  
  
        s.user_id,
```

```

        COUNT(DISTINCT s.trip_id) AS flight_count,

        MAX(f.seats) AS max_seats

FROM

    sessions s

JOIN

    flights f ON s.trip_id = f.trip_id

GROUP BY

    s.user_id

),

-- Step 4: Calculate statistics for flight_count and max_seats
flight_seat_stats AS (

    SELECT

        MIN(flight_count) AS min_flight_count,

        MAX(flight_count) AS max_flight_count,

        AVG(flight_count) AS mean_flight_count,

        PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY flight_count) AS
median_flight_count,

        MIN(max_seats) AS min_max_seats,

        MAX(max_seats) AS max_max_seats,

        AVG(max_seats) AS mean_max_seats,

        PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY max_seats) AS
median_max_seats

```

FROM

user_flight_data

	min_flight_count	max_flight_count	mean_flight_count	median_flight_count	min_max_seats	max_max_seats	mean_max_seats	median_max_seats
1	1	11	2.2535414626744652	2	0	12	1.5054553343022222	1

- **Criteria:** The "Solo Traveler" category is defined by users who book only one seat per flight and have taken at least one flight.
- **Perk:** 100 loyalty point

Max Seats (Median Analysis):

- **Median Max Seats: 1**
- This indicates that half of the users book one seat per flight, which aligns with the definition of a solo traveler. The median value represents the central tendency of the data, showing that booking a single seat is the most common behavior.
- **Flight Count:**
 - **Minimum Flight Count: 1**
 - **Median Flight Count: 2**
 - **Average Flight Count: 2.253**
 - The minimum and median flight counts indicate that users with at least one flight are quite common. The median and average values suggest that while many users fly infrequently, solo travelers still make at least one flight, confirming that they are active but travel alone.
- **Chosen Criteria:**
 - **One Seat per Flight:** This is based on the median max seats value, which is 1. It represents users who consistently book a single seat, defining them as solo travelers.
 - **At Least One Flight:** Ensures the user is an active traveler who has used the service at least once.

Perk for Solo Travelers:

Perk: 100 Loyalty Points

Justification:

1. **Value to Solo Travelers:**

- **Encouragement and Reward:** Awarding **100 loyalty points** to solo travelers provides an incentive for their continued use of the service. It acknowledges their loyalty and encourages them to keep booking flights, even if they are solo travelers.

2. **Alignment with Travel Behavior:**

- **Solo Travel Rewards:** Solo travelers typically do not benefit from perks designed for group travel. Therefore, loyalty points serve as a neutral and flexible reward that can be used for various purposes, such as future flights or upgrades, aligning with their individual travel needs.

3. **Fostering Loyalty:**

- **Retention Strategy:** Offering loyalty points is a common strategy to enhance customer retention. For solo travelers who might not frequently travel with others, these points provide a tangible benefit that encourages ongoing engagement with the travel service.

Summary:

- **Criteria for Solo Travelers:** Users who book **one seat per flight** and have taken **at least one flight**. This definition is supported by the median value of max seats (1), indicating the most common behavior among users.
- **Perk: 100 loyalty points.** This perk is well-suited for solo travelers, offering them a reward that is flexible and can enhance their travel experience without the need for group-specific benefits. It acknowledges their loyalty and encourages further use of the service.