SQL PROJECT

AIRPORT DATA ANALYSIS

OBJECTIVE

The SQL project will analyze the Airport Data to identify the key factors such as airport operations by examining total passenger numbers and trends per route, average passengers per flight, flight frequency. It will compare passenger numbers across origin cities, access available seats, and identify popular destination airports. Additionally, the project will explore correlations between city population and passenger counts, as well as the impact of population size on flight frequency.

Problem Statement 1 :

- The objective is to calculate the total number of passengers for each pair of origin and destination airports.

- This will provide insights into travel patterns between specific airport pairs, helping to identify the most frequented routes and enhance strategic planning for airline operations.

```
SELECT
    Origin_airport,
    Destination_airport,
    SUM(Passengers) AS Total_Passengers
FROM
    airports2
GROUP BY Origin_airport , Destination_airport
ORDER BY Origin_airport , Destination_airport;
```

Problem Statement 2 :

- Here the goal is to calculate the average seat utilization for each flight by dividing the number of passengers by the total number of seats available.

- The results will be sorted in descending order based on utilization percentage.

- This will help identify flights with the highest and lowest seat occupancy, providing valuable insights for optimizing flight capacity and enhancing

```sql
SELECT
    Origin_airport,
    Destination_airport,
    AVG(CAST(Passengers AS FLOAT) / NULLIF(Seats, 0)) * 100 AS Average_Seat_Utilization
FROM
    airports2
GROUP BY
    Origin_airport,
    Destination_airport
ORDER BY
    Average_Seat_Utilization DESC;
```

Problem Statement 3 :

- The aim is to determine the top 5 origin and destination airport pairs that have the highest total passenger volume.

- This analysis will reveal the most frequented travel routes, allowing airlines to optimize resource allocation and enhance service offerings based on passenger demand trends

```
SELECT
    Origin_airport,
    Destination_airport,
    SUM(Passengers) AS Total_Passengers
FROM
    airports2
GROUP BY
    Origin_airport,
    Destination_airport
ORDER BY
    Total_Passengers DESC
LIMIT 5;
```

Problem Statement 4 :

- The objective is to calculate the total number of flights and passengers departing from each origin city.

- This will provide insights into the activity levels at various origin cities, helping to identify key hubs and inform strategic decisions regarding flight operations and capacity management.

```
SELECT
    Origin_city,
    COUNT(Flights) AS Total_Flights,
    SUM(Passengers) AS Total_Passengers
FROM
    airports2
GROUP BY
    Origin_city
ORDER BY
    Origin_city;
```

Problem Statement 5 :

- The aim is to calculate the total distance flown by flights originating from each airport.

- This analysis will offer insights into the overall travel patterns and operational reach of each airport, helping to evaluate their significance in the network and inform future route planning decisions.

```
SELECT
    Origin_airport,
    SUM(Distance) AS Total_Distance
FROM
    airports2
GROUP BY
    Origin_airport
ORDER BY
    Origin_airport;
```

Problem Statement 6 :

- The objective is to group flights by month and year using the Fly_date column to calculate the number of flights, total passengers, and average distance traveled per month.

- This analysis will provide a clearer understanding of seasonal trends and operational performance over time, enabling better strategic planning for airline operations.

```
SELECT
    YEAR(Fly_date) AS Year,
    MONTH(Fly_date) AS Month,
    COUNT(Flights) AS Total_Flights,
    SUM(Passengers) AS Total_Passengers,
    AVG(Distance) AS Avg_Distance
FROM
    airports2
GROUP BY
    YEAR(Fly_date),
    MONTH(Fly_date)
ORDER BY
    Year,
    Month;
```

Problem Statement 7 :

- The goal is to calculate the passenger-to-seats ratio for each origin and destination route and filter the results to display only those routes where this ratio is less than 0.5.

- This will help identify underutilized routes, enabling airlines to make informed decisions about capacity management and potential route adjustments.

```sql
SELECT
    Origin_airport,
    Destination_airport,
    SUM(Passengers) AS Total_Passengers,
    SUM(Seats) AS Total_Seats,
    (SUM(Passengers) * 1.0 / NULLIF(SUM(Seats), 0)) AS Passenger_to_Seats_Ratio
FROM
    airports2
GROUP BY
    Origin_airport,
    Destination_airport
HAVING
    (SUM(Passengers) * 1.0 / NULLIF(SUM(Seats), 0)) < 0.5
ORDER BY
    Passenger_to_Seats_Ratio;
```

Problem Statement 8 :

- The aim is to determine the top 3 origin airports with the highest frequency of flights.

- It will highlight the most active airports in terms of flight operations, providing valuable insights for airlines and stakeholders to optimize scheduling and improve service offerings at these critical locations.

```
SELECT
    Origin_airport,
    COUNT(Flights) AS Total_Flights
FROM
    airports2
GROUP BY
    Origin_airport
ORDER BY
    Total_Flights DESC
LIMIT 3;
```

Problem Statement 9 :

- The objective is to identify the city (excluding Bend, OR) that sends the most flights and passengers to Bend, OR.

- This will reveal key contributors to passenger traffic at Bend, OR, helping airlines and travel authorities understand demand patterns and enhance connectivity from popular originating cities.

```sql
SELECT
    Origin_city,
    COUNT(Flights) AS Total_Flights,
    SUM(Passengers) AS Total_Passengers
FROM
    airports2
WHERE
    Destination_city = 'Bend, OR' AND
    Origin_city <> 'Bend, OR'
GROUP BY
    Origin_city
ORDER BY
    Total_Flights DESC,
    Total_Passengers DESC
LIMIT 3;
```

Problem Statement 10 :

- The aim is to identify the longest flight route in terms of distance traveled, including both the origin and destination airports.

- This will provide insights into the most extensive travel connections, helping airlines assess operational challenges and opportunities for long-haul service planning.

```
SELECT
    Origin_airport,
    Destination_airport,
    MAX(Distance) AS Longest_Distance
FROM
    airports2
GROUP BY
    Origin_airport,
    Destination_airport
ORDER BY
    Longest_Distance DESC
LIMIT 1;
```

CHALLENGING QUESTIONS

PROBLEM STATEMENT FROM 11 TO 20

Problem Statement 11 :

- The objective is to determine the most and least busy months by flight count across multiple years
- This analysis will provide insights into seasonal trends in air travel.
- Helping airlines and stakeholders understand peak and off-peak periods for better operational planning and resource allocation.

```sql
WITH Monthly_Flights AS (
    SELECT
        MONTH(Fly_date) AS Month,
        COUNT(Flights) AS Total_Flights
    FROM
        airports2
    GROUP BY
        MONTH(Fly_date)
)

SELECT
    Month,
    Total_Flights,
    CASE
        WHEN Total_Flights = (SELECT MAX(Total_Flights) FROM Monthly_Flights) THEN 'Most Busy'
        WHEN Total_Flights = (SELECT MIN(Total_Flights) FROM Monthly_Flights) THEN 'Least Busy'
        ELSE NULL
    END AS Month_Status
FROM
    Monthly_Flights
WHERE
    Total_Flights = (SELECT MAX(Total_Flights) FROM Monthly_Flights)
    OR Total_Flights = (SELECT MIN(Total_Flights) FROM Monthly_Flights);
```

Problem Statement 12 :

- The aim is to calculate the year-over-year percentage growth in the total number of passengers for each origin and destination airport pair.

```sql
WITH Passenger_Summary AS (
    SELECT
        Origin_airport,
        Destination_airport,
        YEAR(Fly_date) AS Year,
        SUM(Passengers) AS Total_Passengers
    FROM
        airports2
    GROUP BY
        Origin_airport,
        Destination_airport,
        YEAR(Fly_date)
),

Passenger_Growth AS (
    SELECT
        Origin_airport,
        Destination_airport,
        Year,
        Total_Passengers,
        LAG(Total_Passengers) OVER (PARTITION BY Origin_airport, Destination_airport ORDER BY Year) AS Previous_Year_Passengers
    FROM
        Passenger_Summary
)
```

Problem Statement 12 :

- This analysis will help identify trends in passenger traffic over time

- Providing valuable insights for airlines to make informed decisions about route development and capacity management based on demand fluctuations.

```sql
SELECT
    Origin_airport,
    Destination_airport,
    Year,
    Total_Passengers,
    CASE
        WHEN Previous_Year_Passengers IS NOT NULL THEN
            ((Total_Passengers - Previous_Year_Passengers) * 100.0 / NULLIF(Previous_Year_Passengers, 0))
        ELSE NULL
    END AS Growth_Percentage
FROM
    Passenger_Growth
ORDER BY
    Origin_airport,
    Destination_airport,
    Year;
```

Problem Statement 13 :

- The objective is to identify routes (from origin to destination) that have demonstrated consistent year-over-year growth in the number of flights.

```
WITH Flight_Summary AS (
    SELECT
        Origin_airport,
        Destination_airport,
        YEAR(Fly_date) AS Year,
        COUNT(Flights) AS Total_Flights
    FROM
        airports2
    GROUP BY
        Origin_airport,
        Destination_airport,
        YEAR(Fly_date)
),

Flight_Growth AS (
    SELECT
        Origin_airport,
        Destination_airport,
        Year,
        Total_Flights,
        LAG(Total_Flights) OVER (PARTITION BY Origin_airport, Destination_airport ORDER BY Year) AS Previous_Year_Flights
    FROM
        Flight_Summary
),
```

Problem Statement 13 :

```sql
Growth_Rates AS (
    SELECT
        Origin_airport,
        Destination_airport,
        Year,
        Total_Flights,
        CASE
            WHEN Previous_Year_Flights IS NOT NULL AND Previous_Year_Flights > 0 THEN
                ((Total_Flights - Previous_Year_Flights) * 100.0 / Previous_Year_Flights)
            ELSE NULL
        END AS Growth_Rate,
        CASE
            WHEN Previous_Year_Flights IS NOT NULL AND Total_Flights > Previous_Year_Flights THEN 1
            ELSE 0
        END AS Growth_Indicator
    FROM
        Flight_Growth
)
```

```sql
-- Final query to identify routes with consistent growth and their growth rate
SELECT
    Origin_airport,
    Destination_airport,
    MIN(Growth_Rate) AS Minimum_Growth_Rate,
    MAX(Growth_Rate) AS Maximum_Growth_Rate
FROM
    Growth_Rates
WHERE
    Growth_Indicator = 1
GROUP BY
    Origin_airport,
    Destination_airport
HAVING
    MIN(Growth_Indicator) = 1
ORDER BY
    Origin_airport,
    Destination_airport;
```

This will help airlines understand which routes have not only grown consistently but also the magnitude of that growth in terms of percentage.

also it will highlight successful routes, providing insights for airlines to strengthen their operational strategies and consider potential expansions based on sustained demand trends.

Problem Statement 14 :

- The aim is to determine the top 3 origin airports with the highest weighted passenger-to-seats utilization ratio, considering the total number of flights for weighting.

```sql
WITH Utilization_Ratio AS (
    -- Step 1: Calculate the passenger-to-seats ratio for each flight
    SELECT
        Origin_airport,
        SUM(Passengers) AS Total_Passengers,
        SUM(Seats) AS Total_Seats,
        COUNT(Flights) AS Total_Flights,
        SUM(Passengers) * 1.0 / SUM(Seats) AS Passenger_Seat_Ratio
    FROM
        airports2
    GROUP BY
        Origin_airport
),

Weighted_Utilization AS (
    -- Step 2: Calculate the weighted utilization by flights for each origin airport
    SELECT
        Origin_airport,
        Total_Passengers,
        Total_Seats,
        Total_Flights,
        Passenger_Seat_Ratio,
        -- Weight the passenger-to-seat ratio by the total number of flights
        (Passenger_Seat_Ratio * Total_Flights) / SUM(Total_Flights) OVER () AS Weighted_Utilization
    FROM
        Utilization_Ratio
)
```

Problem Statement 14 :

- It will highlight the top 3 origin airports that not only have good passenger-to-seat ratios but also perform well when the total number of flights is considered. It gives a more balanced view of operational efficiency by considering both the ratio and flight volume.

```
-- Step 3: Select the top 3 airports by weighted utilization
SELECT
    Origin_airport,
    Total_Passengers,
    Total_Seats,
    Total_Flights,
    Weighted_Utilization
FROM
    Weighted_Utilization
ORDER BY
    Weighted_Utilization DESC
LIMIT 3;
```

Problem Statement 15 :

- The objective is to identify the peak traffic month for each origin city based on the highest number of passengers, including any ties where multiple months have the same passenger count.

```sql
WITH Monthly_Passenger_Count AS (
    SELECT
        Origin_city,
        YEAR(Fly_date) AS Year,
        MONTH(Fly_date) AS Month,
        SUM(Passengers) AS Total_Passengers  -- Handling NULLs and non-integer values
    FROM
        airports2
    GROUP BY
        Origin_city,
        YEAR(Fly_date),
        MONTH(Fly_date)
),

Max_Passengers_Per_City AS (
    SELECT
        Origin_city,
        MAX(Total_Passengers) AS Peak_Passengers
    FROM
        Monthly_Passenger_Count
    GROUP BY
        Origin_city
)
```

Problem Statement 15 :

- This will help reveal seasonal travel patterns specific to each city,enabling airlines to tailor their services and marketing strategies to meet demand effectively.

```sql
SELECT
    mpc.Origin_city,
    mpc.Year,
    mpc.Month,
    mpc.Total_Passengers
FROM
    Monthly_Passenger_Count mpc
JOIN
    Max_Passengers_Per_City mp ON mpc.Origin_city = mp.Origin_city
                               AND mpc.Total_Passengers = mp.Peak_Passengers
ORDER BY
    mpc.Origin_city,
    mpc.Year,
    mpc.Month;
```

Problem Statement 16 :

- The aim is to identify the routes (origin-destination pairs) that have experienced the largest decline in passenger numbers year-over-year.

```sql
WITH Yearly_Passenger_Count AS (
    SELECT
        Origin_airport,
        Destination_airport,
        YEAR(Fly_date) AS Year,
        SUM(Passengers) AS Total_Passengers
    FROM
        airports2
    GROUP BY
        Origin_airport,
        Destination_airport,
        YEAR(Fly_date)
),
```

```sql
Yearly_Decline AS (
    SELECT
        y1.Origin_airport,
        y1.Destination_airport,
        y1.Year AS Year1,
        y1.Total_Passengers AS Passengers_Year1,
        y2.Year AS Year2,
        y2.Total_Passengers AS Passengers_Year2,
        -- Calculate percentage decline: (New - Old) / Old * 100
        ((y2.Total_Passengers - y1.Total_Passengers) / NULLIF(y1.Total_Passengers, 0)) * 100 AS Percentage_Change
    FROM
        Yearly_Passenger_Count y1
    JOIN
        Yearly_Passenger_Count y2
        ON y1.Origin_airport = y2.Origin_airport
        AND y1.Destination_airport = y2.Destination_airport
        AND y2.Year = y1.Year + 1 -- Join consecutive years
)
```

Problem Statement 16 :

- This will help airlines pinpoint routes facing reduced demand, allowing for strategic adjustments in operations, marketing, and service offerings to address the decline effectively.

```sql
SELECT
    Origin_airport,
    Destination_airport,
    Year1,
    Year2,
    Passengers_Year1,
    Passengers_Year2,
    Percentage_Change
FROM
    Yearly_Decline
WHERE
    Percentage_Change < 0 -- Only declining routes
ORDER BY
    Percentage_Change ASC -- Largest decline first
LIMIT 5;
```

Problem Statement 17 :

- The objective is to list all origin and destination airports that had at least 10 flights but maintained an average seat utilization (passengers/seats) of less than 50%.

```
WITH Flight_Stats AS (
    SELECT
        Origin_airport,
        Destination_airport,
        COUNT(Flights) AS Total_Flights,
        SUM(Passengers) AS Total_Passengers,
        SUM(Seats) AS Total_Seats,
        -- Calculate average seat utilization as (Total Passengers / Total Seats)
        (SUM(Passengers) / NULLIF(SUM(Seats), 0)) AS Avg_Seat_Utilization
    FROM
        airports2
    GROUP BY
        Origin_airport, Destination_airport
)
```

Problem Statement 17 :

- It will highlight underperforming routes, allowing airlines to reassess their capacity management strategies and make informed decisions regarding potential service adjustments to optimize seat utilization and improve profitability

```sql
SELECT
    Origin_airport,
    Destination_airport,
    Total_Flights,
    Total_Passengers,
    Total_Seats,
    ROUND(Avg_Seat_Utilization * 100, 2) AS Avg_Seat_Utilization_Percentage
FROM
    Flight_Stats
WHERE
    Total_Flights >= 10 -- At least 10 flights
    AND Avg_Seat_Utilization < 0.5 -- Less than 50% seat utilization
ORDER BY
    Avg_Seat_Utilization_Percentage ASC;
```

Problem Statement 18 :

- The aim is to calculate the average flight distance for each unique city-to-city pair (origin and destination) and identify the routes with the longest average distance.
- This analysis will provide insights into long-haul travel patterns, helping airlines assess operational consideration and potential market opportunities for extended routes.

```sql
WITH Distance_Stats AS (
    SELECT
        Origin_city,
        Destination_city,
        AVG(Distance) AS Avg_Flight_Distance
    FROM
        airports2
    GROUP BY
        Origin_city,
        Destination_city
)

SELECT
    Origin_city,
    Destination_city,
    ROUND(Avg_Flight_Distance, 2) AS Avg_Flight_Distance
FROM
    Distance_Stats
ORDER BY
    Avg_Flight_Distance DESC;  -- Sort by average distance in descending order
```

Problem Statement 19 :

- The objective is to calculate the total number of flights and passengers for each year, along with the percentage growth in both flights and passengers compared to the previous year.

```
WITH Yearly_Summary AS (
    SELECT
        SUBSTR(Fly_date, 7, 4) AS Year,  -- Extracting year from the Fly_date string
        COUNT(Flights) AS Total_Flights,
        SUM(Passengers) AS Total_Passengers
    FROM
        airports2
    GROUP BY
        SUBSTR(Fly_date, 7, 4)
),

Yearly_Growth AS (
    SELECT
        Year,
        Total_Flights,
        Total_Passengers,
        LAG(Total_Flights) OVER (ORDER BY Year) AS Prev_Flights,
        LAG(Total_Passengers) OVER (ORDER BY Year) AS Prev_Passengers
    FROM
        Yearly_Summary
)
```

Problem Statement 19 :

- It will provide a comprehensive overview of annual trends in air travel, enabling airlines and stakeholders to assess growth patterns and make informed strategic decisions for future operations.

```sql
SELECT
    Year,
    Total_Flights,
    Total_Passengers,
    ROUND(((Total_Flights - Prev_Flights) / NULLIF(Prev_Flights, 0) * 100), 2) AS Flight_Growth_Percentage,
    ROUND(((Total_Passengers - Prev_Passengers) / NULLIF(Prev_Passengers, 0) * 100), 2) AS Passenger_Growth_Percentage
FROM
    Yearly_Growth
ORDER BY
    Year;
```

Problem Statement 20 :

- The aim is to identify the top 3 busiest routes (origin destination pairs) based on the total distance flown, weighted by the number of flights.

```
WITH Route_Distance AS (
    SELECT
        Origin_airport,
        Destination_airport,
        SUM(Distance) AS Total_Distance,
        SUM(Flights) AS Total_Flights
    FROM
        airports2
    GROUP BY
        Origin_airport,
        Destination_airport
),

Weighted_Routes AS (
    SELECT
        Origin_airport,
        Destination_airport,
        Total_Distance,
        Total_Flights,
        Total_Distance * Total_Flights AS Weighted_Distance
    FROM
        Route_Distance
)
```

Problem Statement 20 :

- This will highlight the most significant routes in terms of distance and operational activity, providing valuable insights for airlines to optimize their scheduling and resource allocation strategies.

```
SELECT
    Origin_airport,
    Destination_airport,
    Total_Distance,
    Total_Flights,
    Weighted_Distance
FROM
    Weighted_Routes
ORDER BY
    Weighted_Distance DESC
LIMIT 3;  -- To get the top 3 busiest routes
```

Story telling : Identifying Operational Challenges in the Airline Industry

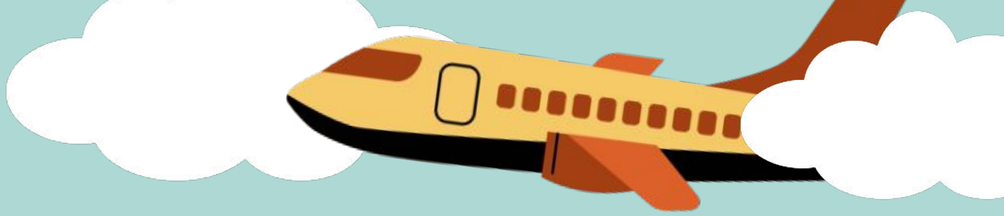
The airline faced several operational difficulties, ranging from declining route performance to underutilized flight capacity. Despite growing passenger demand in some areas, inefficiencies were noted across several routes, with some underperforming. Additionally, seasonal fluctuations weren't being fully leveraged, leading to mismatches between supply and demand.

These problems created a ripple effect:

- Declining profitability on key routes.
- Wasted resources due to low seat utilization.
- Missed opportunities for expansion and market leadership.

This led to the need for a comprehensive data analysis to pinpoint the root causes of these issues and drive actionable insights.

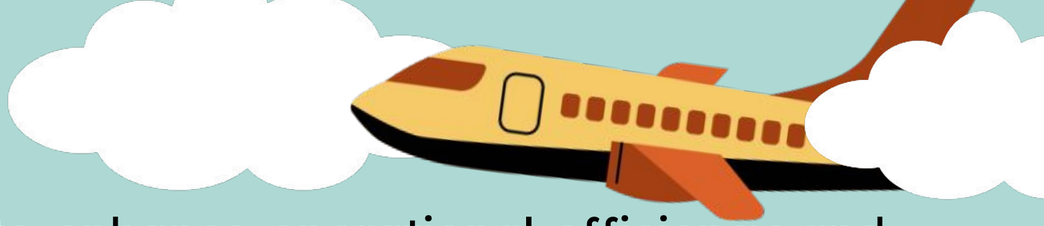
Insights Derived from Data Analysis(SQL)



By analyzing flight data, several critical factors were uncovered that were affecting overall performance:

1. **Inconsistent Route Growth:**
   - Certain routes were thriving while others showed a year-over-year decline in passenger numbers. This indicates that demand forecasting and route planning needed to be realigned.
1. **Low Seat Utilization:**
   - Routes with less than 50% seat utilization were identified, indicating over-capacity on flights. This was a major contributor to operational inefficiencies.
1. **Seasonal Peaks in Demand:**
   - Cities experienced predictable surges in passenger numbers during certain months, but the airline's scheduling wasn’t fully aligned with these peaks, leading to unbalanced resource allocation.

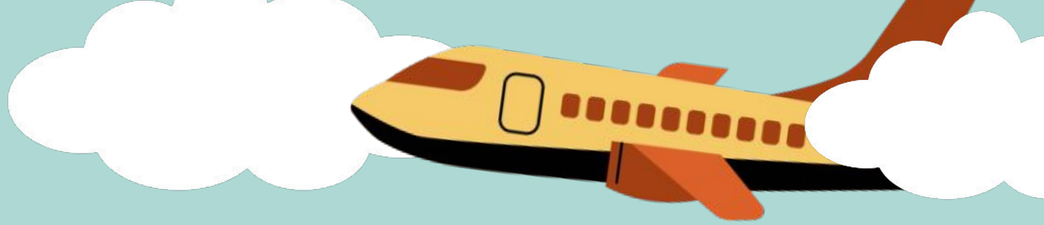
Proposed Strategic Solutions



To address these issues, the following strategies are proposed to enhance operational efficiency and increase profitability:

1. **Route Optimization and Realignment:**
    - Reduce flight frequency or capacity on underperforming routes with declining passengers.
    - Invest in expanding capacity on routes demonstrating consistent growth, as highlighted in the year-over-year trends.
2. **Enhance Seat Utilization:**
    - Realign aircraft sizes to better match passenger demand on low-utilization routes.
    - Implement dynamic pricing and targeted promotions to fill more seats, especially on off-peak flights.
3. **Leverage Seasonal Demand:**
    - Use historical data on peak months to increase flights or larger aircraft during high-demand periods.
    - Tailor marketing strategies to attract passengers during these seasons, ensuring full use of available capacity.
4. **Maximize Long-Distance Route Efficiency:**
    - Focus on routes with the longest average distance and significant passenger volume to ensure

Conclusion and Stretegic Roadmap :



In conclusion, our data-driven approach has highlighted key operational inefficiencies and provided actionable insights to tackle these challenges.

The airline can:

- Enhance profitability by focusing on high-growth routes and optimizing underperforming ones.
- Improve resource allocation through better seat utilization and more efficient capacity management.
- Capitalize on seasonal trends to match service levels with fluctuating passenger demand, maximizing revenue potential during peak months.

By implementing these strategies, the airline can position itself for sustainable growth, reduce operational costs, and offer a more consistent and satisfactory experience for passengers moving forward.

Thank You !