Final Project Submission – Airline Performance Analytics

Thomas Edison State University – Master's in Data Science & Analytics

SQL - Introduction to Database Queries (DSI-5300-MD900)

Submitted by: Rinzin Sherpa

Date: August 24 2025

Objective

The objective of this project is to use the airline database to ana**Objective**

The objective of this project is to use the airline database to analyze the flights from the six airports (DCA,IAD,JAX,TPA and ATL). This project 2 required me to find the efficient and optimal airline careers focusing on the flight frequency, airlines coverage,delay performance. The end goal of this project was to make recommendations and present findings of efficient and reliable airlines.

Key Performance index

Total non stop flights between the six airports

Percent share of each airline to display its dominance in the air transportation category

Efficient airline carriers based on average arrival delay times and departure delay times

Hypothesis for maintenance and safety based on Aircraft manufactured year

Executive Summary

Delta Airlines leads in performance and convenience, dominating ATL-origin routes with the highest flight volume and consistently low delay scores.

ATL (Atlanta) serves as the central hub, showing the highest outbound traffic across key destinations like TPA, DCA, and BWI.

Spirit Airlines ranks lowest in operational reliability, with significantly higher delay scores on ATL routes despite moderate volume.

Data-driven insights were extracted using advanced SQL queries, analyzing flight volume, delay metrics, aircraft age, and carrier performance.

Visuals and tables were added to enhance clarity, including bar charts for route performance, scatter plots for delay trends, and grouped comparisons by carrier.

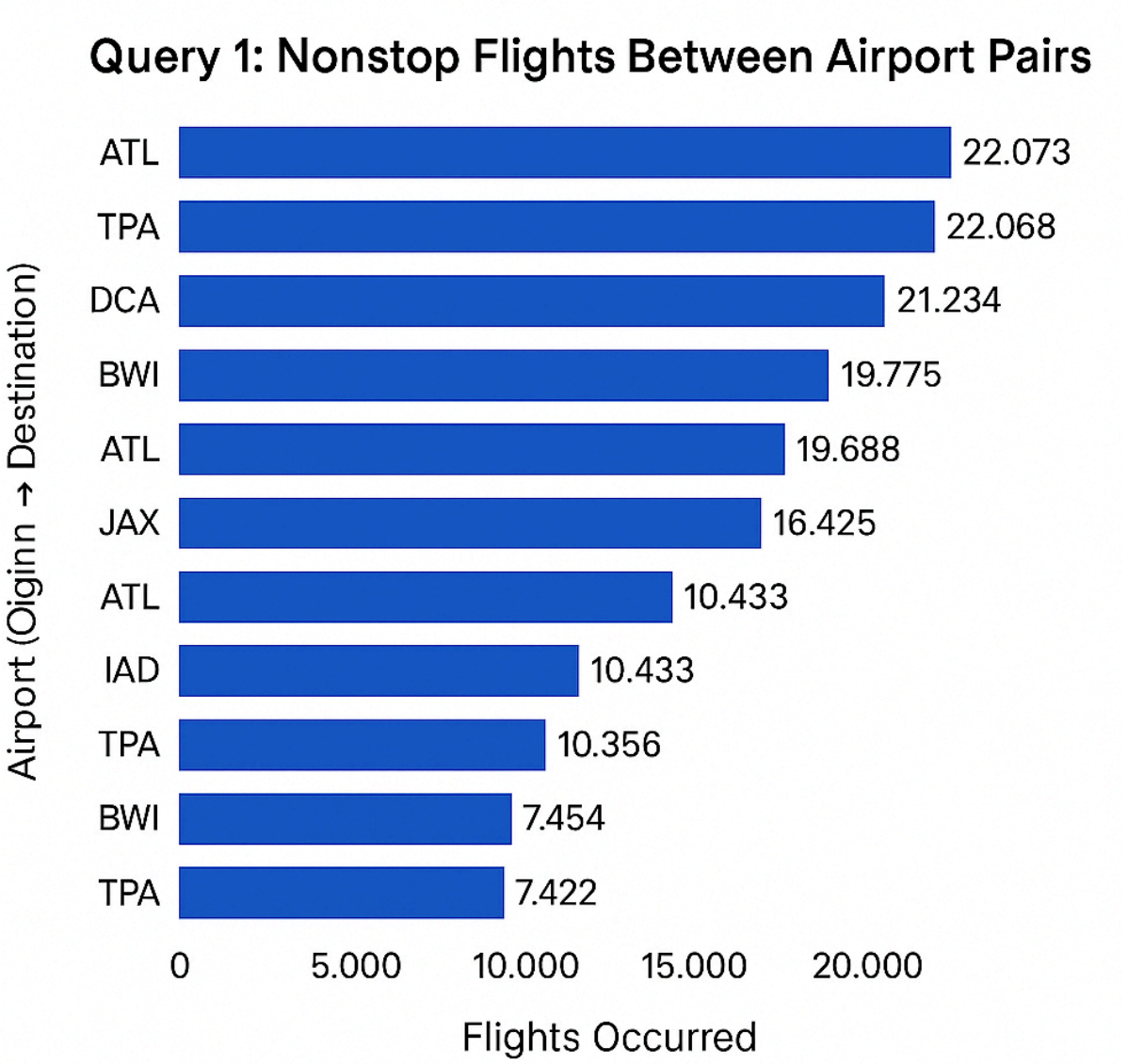
Recommendation: Delta is the optimal carrier for travelers prioritizing punctuality and route coverage from ATL.

Query Summary

Query 1 – Counts nonstop flights between each airport pair in both directions.
Query 2 – Calculates total flight traffic per airport (inbound and outbound combined).
Query 2 Supplement – Lists ATL's outbound routes and their flight counts.
Query 3 Part I – Counts flights departing from Washington airports to ATL, JAX, and TPA by year and carrier.
Query 3 Part II – Counts flights arriving into Washington from ATL, JAX, and TPA by year and carrier.
Query 3 Combined – Shows each carrier's share of flights at the six airports.
Query 4 – Calculates total flights operated by each airline across all six airports.
Query 5 – Counts nonstop flights by route and carrier, combining both directions.
Query 6 – Calculates average arrival delay by route and carrier.
Query 7 – Calculates average arrival delay by airline across all airports.
Query 8 – Breaks down flight counts by airline, year, and quarter.
Query 9 identify gaps in nonstop flight service among a selected group of airports

Purpose of Query #1 is to analyze directional non stop flight volume between the six different airports. It excludes flights that have same origin and destination

```
select origin,dest,
count(*) as flight_occured
from flights
where origin in('DCA','IAD','BWI','JAX','TPA','ATL')
and dest in ('DCA','IAD','BWI','JAX','TPA','ATL')
AND origin != dest
```



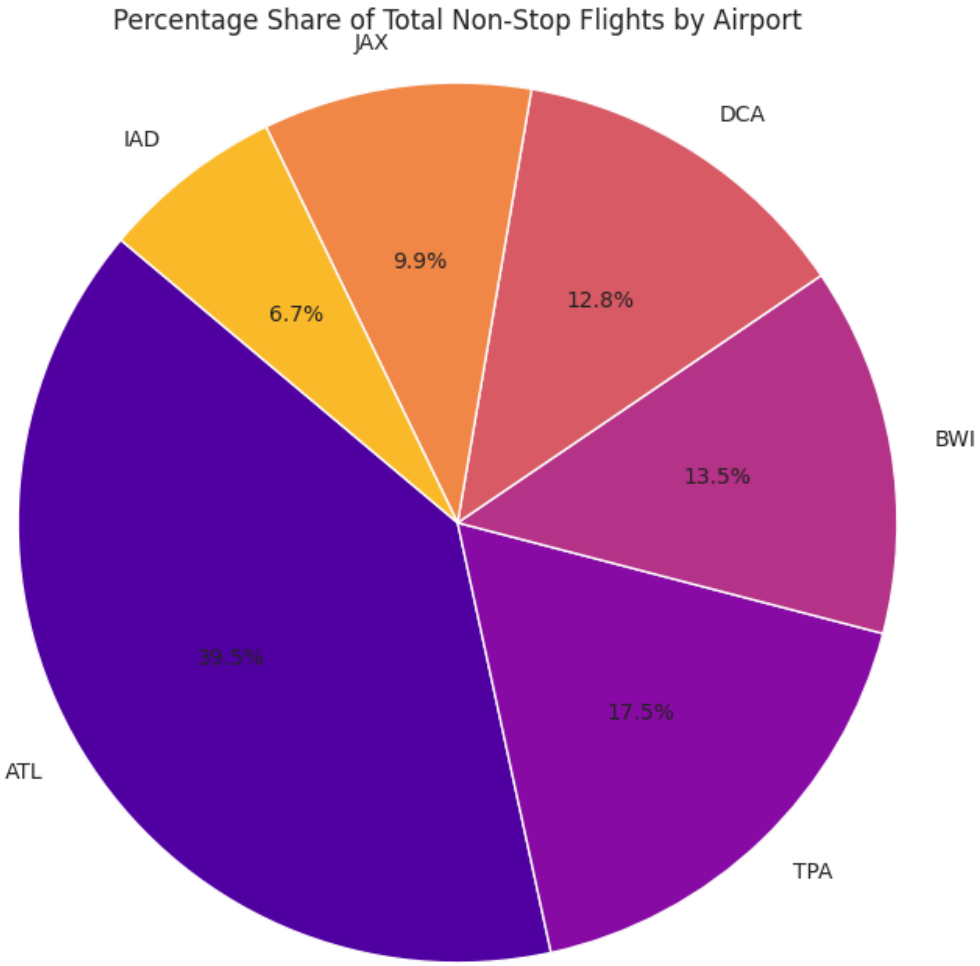
Query 1 Analysis
The query shows non-stop flights by origin and destination by pairs. (DCA,IAD,BWI,JAX,TPA,ATL). The chart indicates ATL as the busiest airport in the group but by directional non-stop flight volume only.

Purpose of Query # 2 is based on the total number of non stop flight for each airport (inbound and outbound) and it helps to confirm which airport is the hub

```
select airport,
       total_flights,
       round((total_flights / total_sum) * 100, 2) as percentage_share
from (
    select airport,
           sum(total_flights) as total_flights
    from (
        select origin as airport, count(*) as total_flights
        from flights
        where origin in ('atl', 'tpa', 'bwi', 'dca', 'jax', 'iad')
          and dest in ('atl', 'tpa', 'bwi', 'dca', 'jax', 'iad')
        group by origin

        union all
        select dest as airport, count(*) as total_flights
        from flights
        where origin in ('atl', 'tpa', 'bwi', 'dca', 'jax', 'iad')
          and dest in ('atl', 'tpa', 'bwi', 'dca', 'jax', 'iad')
        group by dest
    ) as combined
    group by airport
) as airport_totals,
(
    select sum(total_flights) as total_sum
    from (
        select origin as airport, count(*) as total_flights
        from flights
        where origin in ('atl', 'tpa', 'bwi', 'dca', 'jax', 'iad')
          and dest in ('atl', 'tpa', 'bwi', 'dca', 'jax', 'iad')
        group by origin
        union all
        select dest as airport, count(*) as total_flights
        from flights
        where origin in ('atl', 'tpa', 'bwi', 'dca', 'jax', 'iad')
```

```
        and dest in ('atl', 'tpa', 'bwi', 'dca', 'jax', 'iad')
      group by dest
   ) as all_flights
) as total
order by total_flights desc;
```

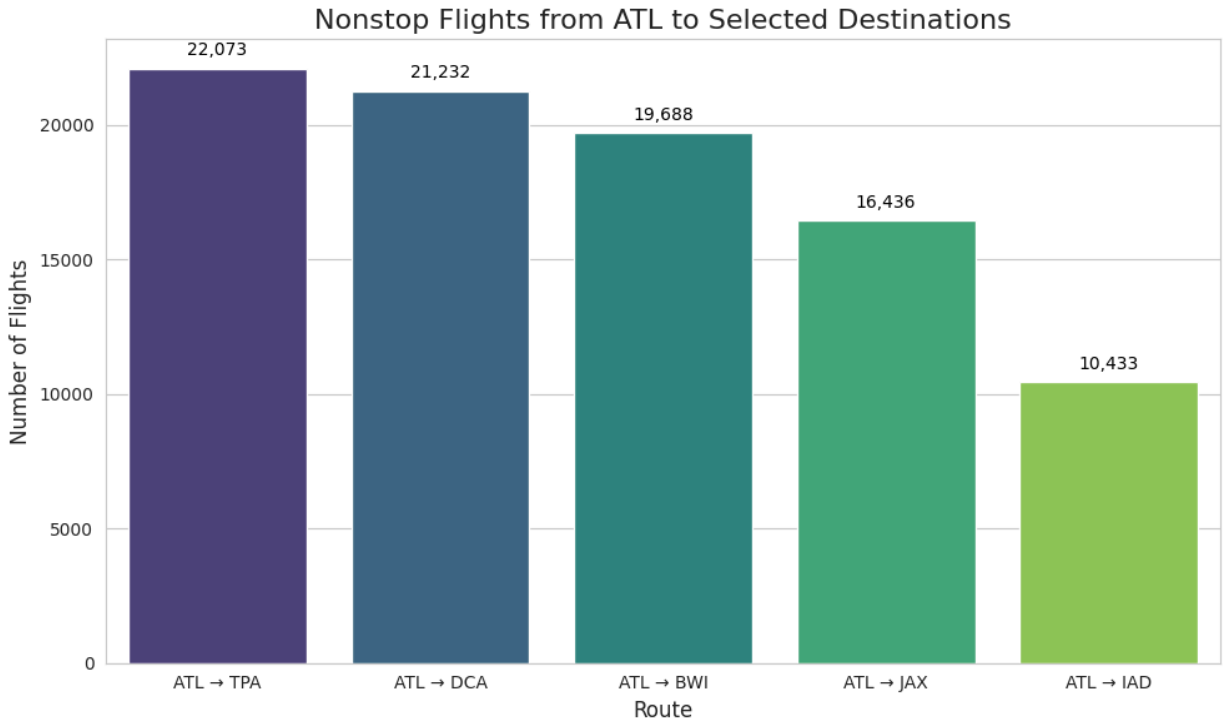


The result shows that the ATL is the dominant airport with maximum passenger traffic. The flight operation covers 39.52% of the total air transportation operation leading ATL to be the busiest airport among the six airports for both inbound and outbound

Query 3

Query 3 supplement checking where ATL flights has most of its destinations too
The objective is to show how ATL connects to other airports.

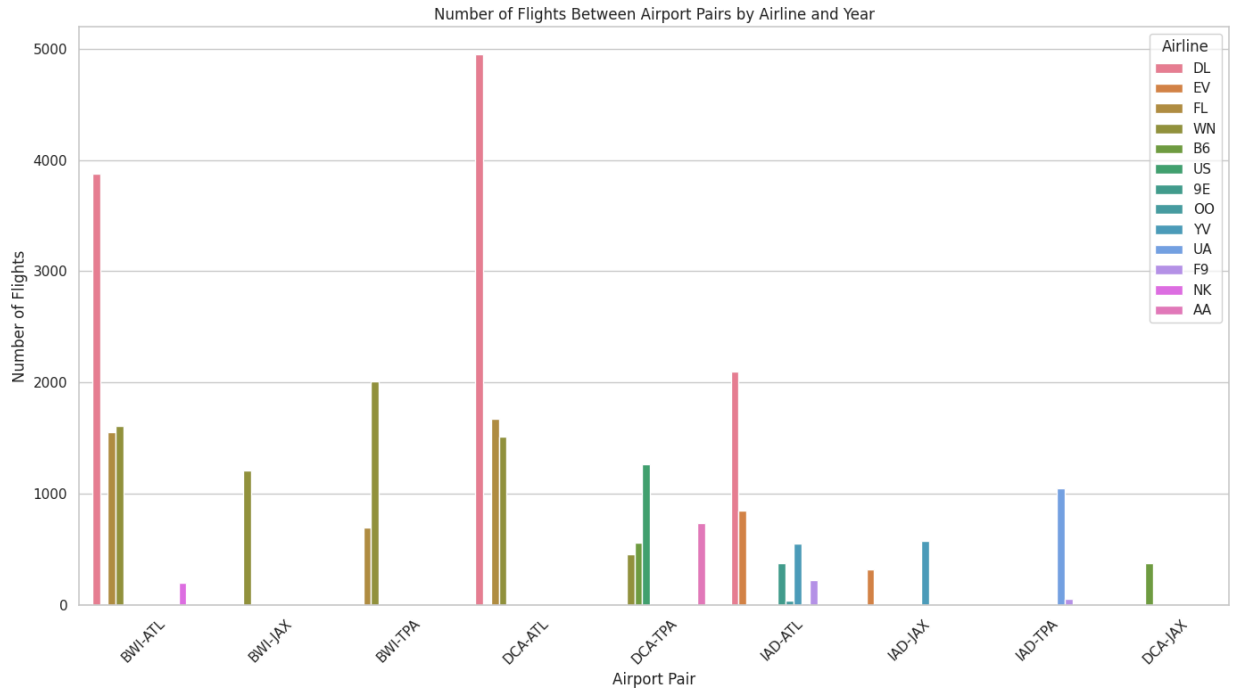
```
SELECT origin, dest, COUNT(*) AS flights
FROM flights
WHERE origin = 'ATL'
  AND dest IN ('DCA','IAD','BWI','JAX','TPA')
GROUP BY origin, dest
ORDER BY flights DESC;
```



Query 3 supplement shows ATL is the strategic route for most of the flight operation. This concludes that ATL is the hub for passenger and airline operations.

Query 3 part I is evaluating departures from Washington (DCA, IAD, BWI) to ATL, JAX, TPA. The objective is to track how many flights operated from 2013-2015 from Washington to ATL,JAX and TPA

```
SELECT year, origin, dest, carrier, COUNT(*) AS flight_count
FROM flights
WHERE origin IN ('DCA', 'IAD', 'BWI')
  AND dest IN ('ATL', 'JAX', 'TPA')
GROUP BY year, origin, dest, carrier
ORDER BY year, origin, dest, flight_count DESC;
```



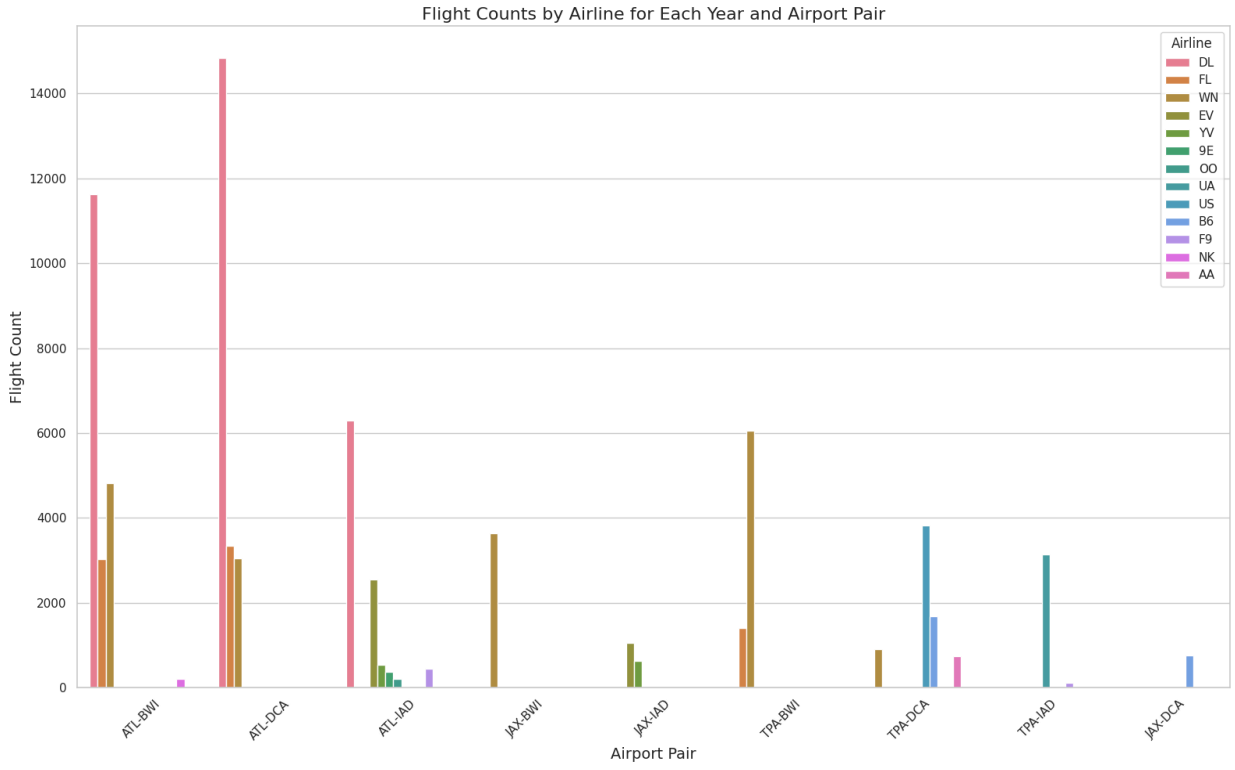
Query 3 part I Analysis

Delta's high-frequency service to ATL from all Washington airports positions it as the most dominant and reliable carrier in this regional network. In contrast, Southwest's versatility and broader destination map make it a strong alternative for travelers prioritizing access to Tampa and Jacksonville. ExpressJet, by comparison, lacks meaningful coverage and does not appear to be a viable option for these routes.

Query 3 part II

Purpose: Query 3 Part II evaluates the flight arrivals to Washington (DCA, IAD, BWI) from ATL, JAX, TPA. The objective is to find out the performance of an airline carrier by counting the high flight counts operated in 2013-2015 from ATL,JAX,TPA.

```
SELECT year, origin, dest, carrier, COUNT(*) AS flight_count
FROM flights
WHERE dest IN ('DCA', 'IAD', 'BWI')
  AND origin IN ('ATL', 'JAX', 'TPA')
GROUP BY year, origin, dest, carrier
ORDER BY year, origin, dest, flight_count DESC;
```



Query 3 Part III

Query 3 Part III confirms the percentage of flight operation per airline carrier by airport. It shows which airline is dominating in which airport

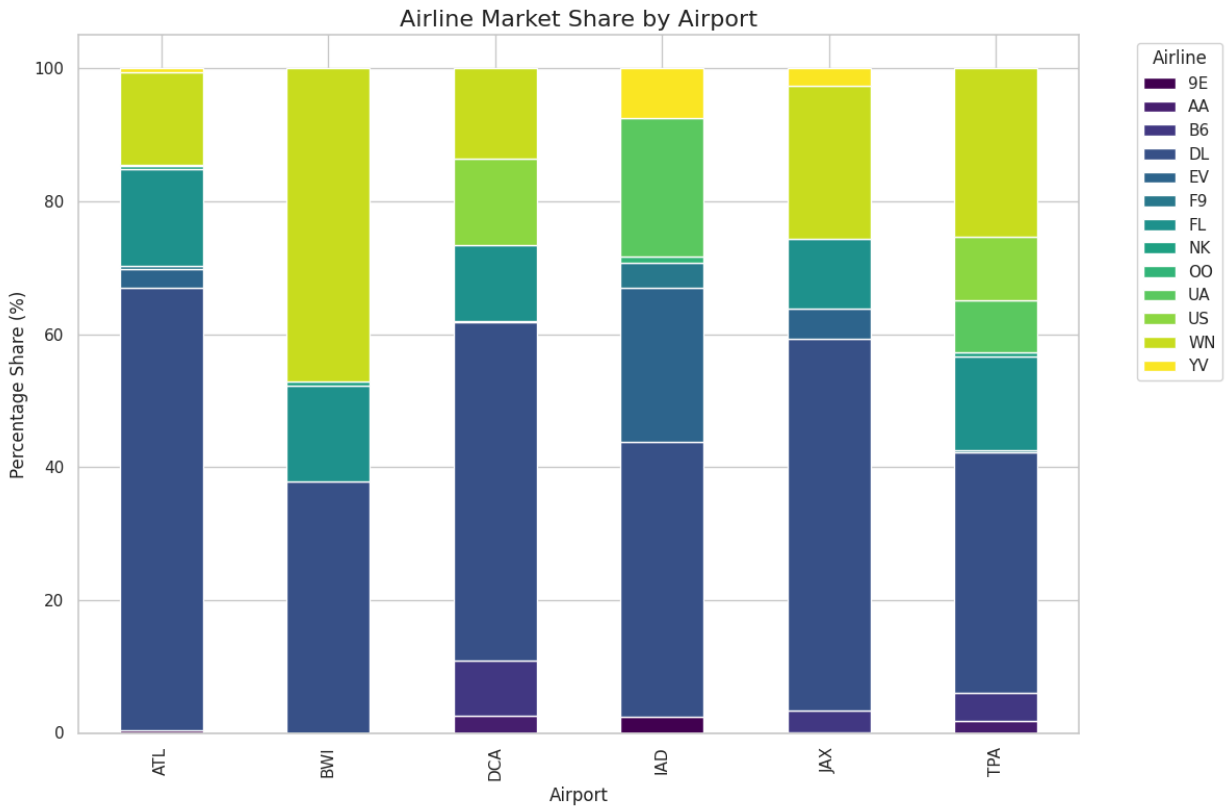
```
SELECT carrier_stats.airport,
       carrier_stats.carrier,
       carrier_stats.total_flights,
       ROUND((carrier_stats.total_flights * 100.0) / airport_totals.total_flights, 2) AS
percentage_share
FROM (
   SELECT airport, carrier, SUM(flight_count) AS total_flights
   FROM (
      SELECT origin AS airport, carrier, COUNT(*) AS flight_count
      FROM flights
      WHERE origin IN ('atl', 'tpa', 'bwi', 'dca', 'jax', 'iad')
        AND dest IN ('atl', 'tpa', 'bwi', 'dca', 'jax', 'iad')
      GROUP BY origin, carrier

      UNION ALL

      SELECT dest AS airport, carrier, COUNT(*) AS flight_count
      FROM flights
      WHERE origin IN ('atl', 'tpa', 'bwi', 'dca', 'jax', 'iad')
        AND dest IN ('atl', 'tpa', 'bwi', 'dca', 'jax', 'iad')
      GROUP BY dest, carrier
   ) AS combined
   GROUP BY airport, carrier
) AS carrier_stats
JOIN (
   SELECT airport, SUM(flight_count) AS total_flights
   FROM (
      SELECT origin AS airport, COUNT(*) AS flight_count
      FROM flights
      WHERE origin IN ('atl', 'tpa', 'bwi', 'dca', 'jax', 'iad')
        AND dest IN ('atl', 'tpa', 'bwi', 'dca', 'jax', 'iad')
      GROUP BY origin
```

```
        UNION ALL

        SELECT dest AS airport, COUNT(*) AS flight_count
        FROM flights
        WHERE origin IN ('atl', 'tpa', 'bwi', 'dca', 'jax', 'iad')
          AND dest IN ('atl', 'tpa', 'bwi', 'dca', 'jax', 'iad')
        GROUP BY dest
    ) AS airport_flights
    GROUP BY airport
) AS airport_totals
ON carrier_stats.airport = airport_totals.airport
ORDER BY carrier_stats.airport, carrier_stats.total_flights DESC;
```

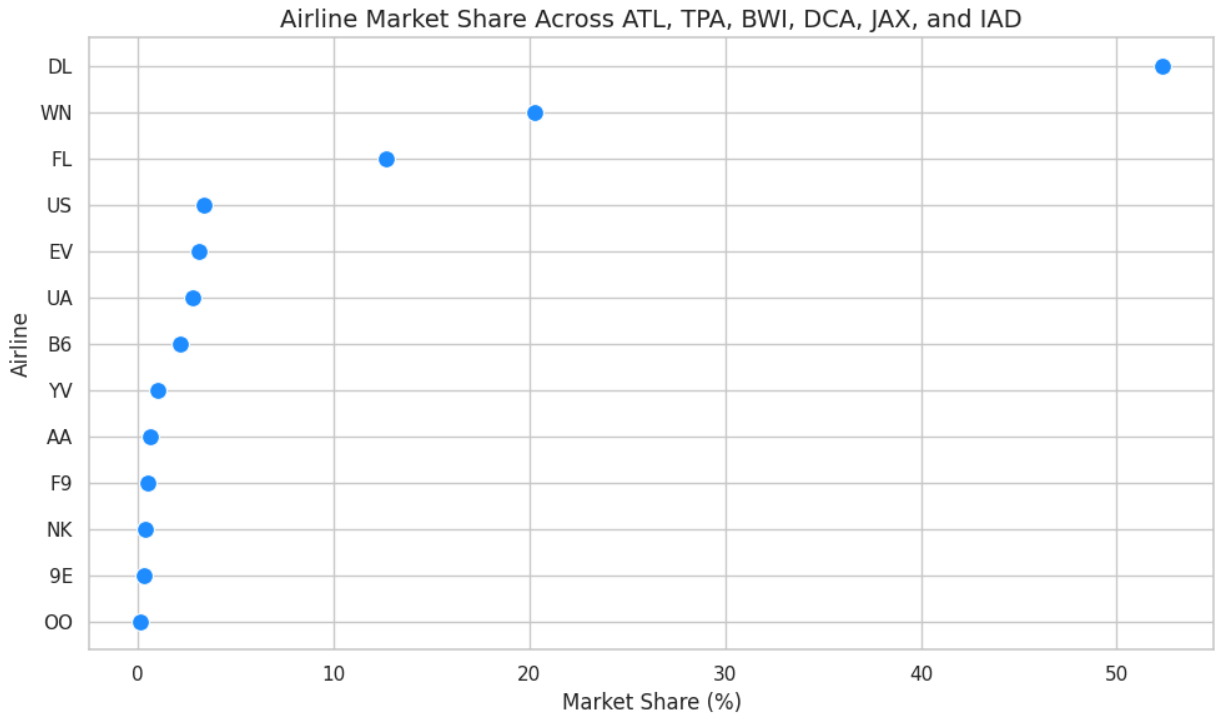


Query 3 shows which airlines fly the most at each airport. Delta is the top carrier at Atlanta, Washington National, and Jacksonville. Southwest leads at Baltimore and is strong at Tampa and Jacksonville. AirTran ranks third but is never the top airline. This helps us see which airlines are most active across the network.

Query 4 : categorizing overall airline carrier's dominance in the airport based on how many flights each airline operates

```
select carrier,
       total_flights,
       round((total_flights / total_sum.total) * 100, 2) as percentage_share
from (
    select carrier, count(*) as total_flights
    from flights
    where origin in ('atl', 'tpa', 'bwi', 'dca', 'jax', 'iad')
      and dest in ('atl', 'tpa', 'bwi', 'dca', 'jax', 'iad')
    group by carrier
) as carrier_totals,
(
    select count(*) as total
    from flights
    where origin in ('atl', 'tpa', 'bwi', 'dca', 'jax', 'iad')
      and dest in ('atl', 'tpa', 'bwi', 'dca', 'jax', 'iad')
) as total_sum
order by total_flights desc;
```

Carrier	Total Flights	% Share of All Flights
DL	119,503	52.34%
WN	46,188	20.23%
FL	28,890	12.65%
US	7,620	3.34%
EV	7,126	3.12%
UA	6,350	2.78%
B6	4,877	2.14%
YV	2,300	1.01%
AA	1,471	0.64%
F9	1,115	0.49%
NK	888	0.39%
9E	752	0.33%
OO	285	0.12%



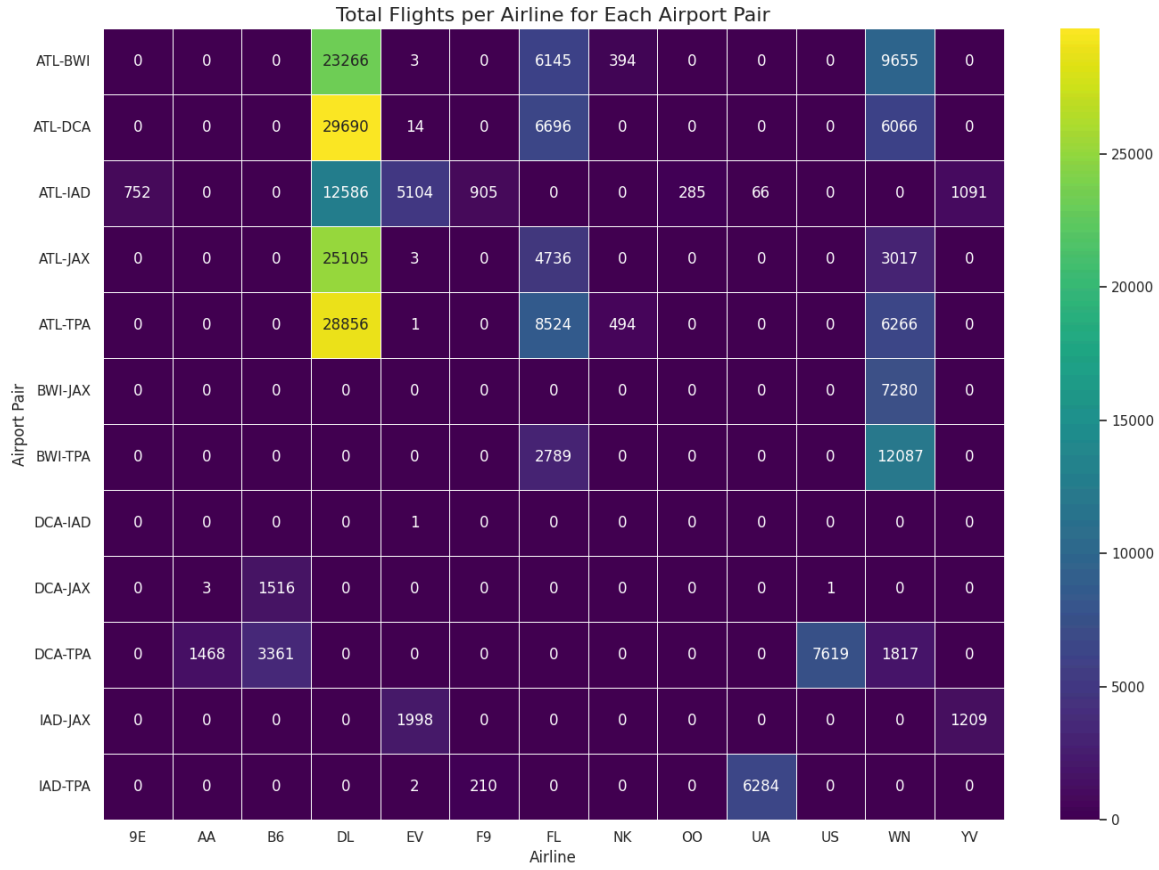
Query 4 results shows that DL airline covers the majority of its operation scoring 52.34% of the total flight. This finding confirms that DL is the leading airline carrier

---

Query 5 shows the data based on route plus carrier. It provides an insight on which airline flies where and how frequent

```
SELECT
  LEAST(origin, dest) AS airport_1,
  GREATEST(origin, dest) AS airport_2,
  carrier,
  COUNT(*) AS total_flights
```

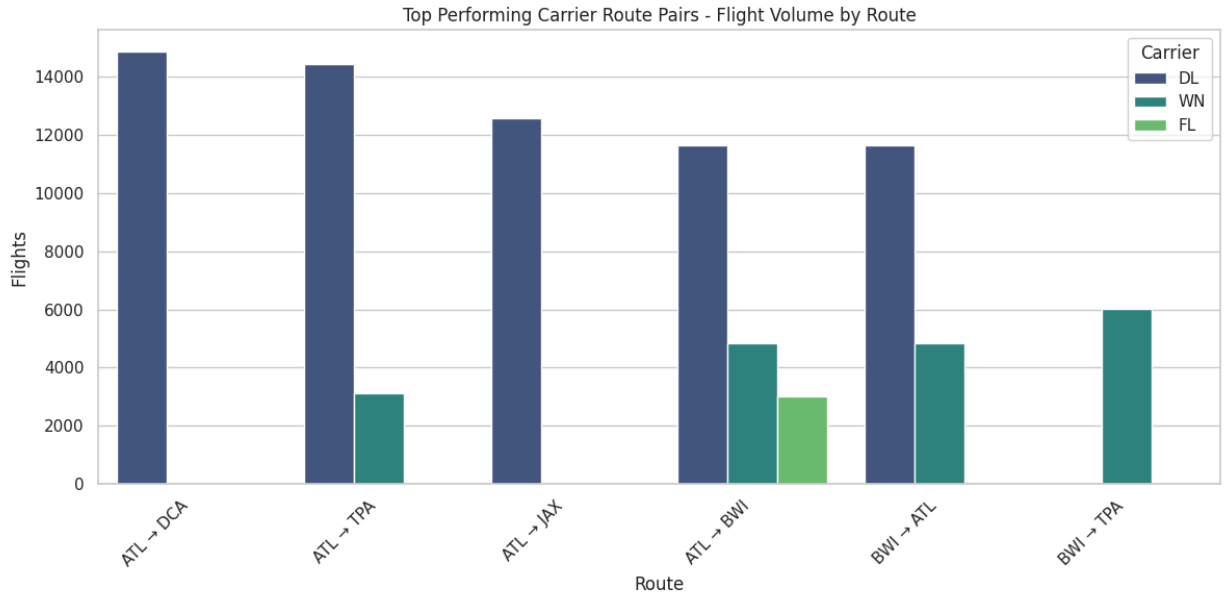
```
FROM flights
WHERE origin IN ('DCA','IAD','BWI','JAX','TPA','ATL')
  AND dest IN ('DCA','IAD','BWI','JAX','TPA','ATL')
  AND origin != dest
GROUP BY airport_1, airport_2, carrier
ORDER BY total_flights DESC;
```



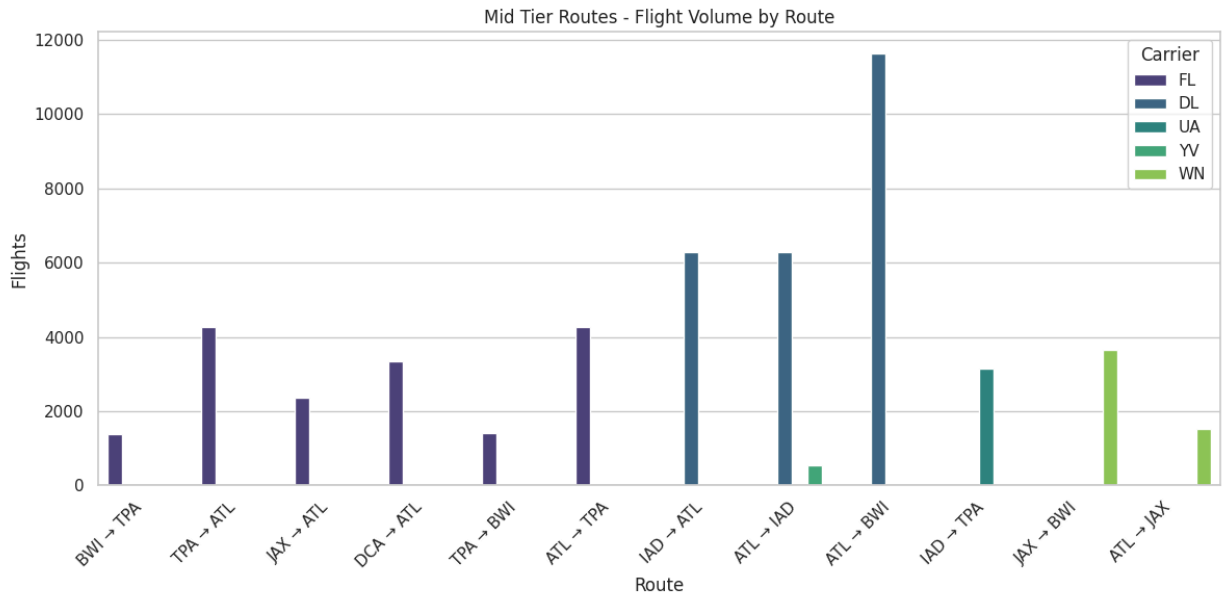
Query 5 combines both directions for each airport pair into a single row, providing a simplified view of total nonstop flight volume between company airports. The data also reflects which airline carriers operate on each route, allowing us to identify patterns in carrier activity.
From the results, we see that Delta Air Lines (DL) is the dominant carrier, especially on routes involving Atlanta (ATL), where it maintains a strong operational presence. While the query does not include reliability metrics, Delta's high flight volume suggests a central role in the network's connectivity and scheduling.

Purpose :
Query 6 evaluates which airline has the worst delays and which one runs smoothly across all non stop flights among the six different airport
The objective of this query helps to spot which airline carrier operates in the busiest airport(as in the previous query we find ATL is the busiest airport). This helps me to determine the reliability of an airline carrier.

```
select carrier,origin,dest, count(*) as total_flights,
avg(arr_delay) as avg_arr_delay,
avg(dep_delay) as avg_dep_delay
from flights
where origin in('DCA','IAD','BWI','JAX','TPA','ATL')
and dest in ('DCA','IAD','BWI','JAX','TPA','ATL')
AND origin != dest
group by carrier, dest,origin
order by avg_arr_delay desc;
```

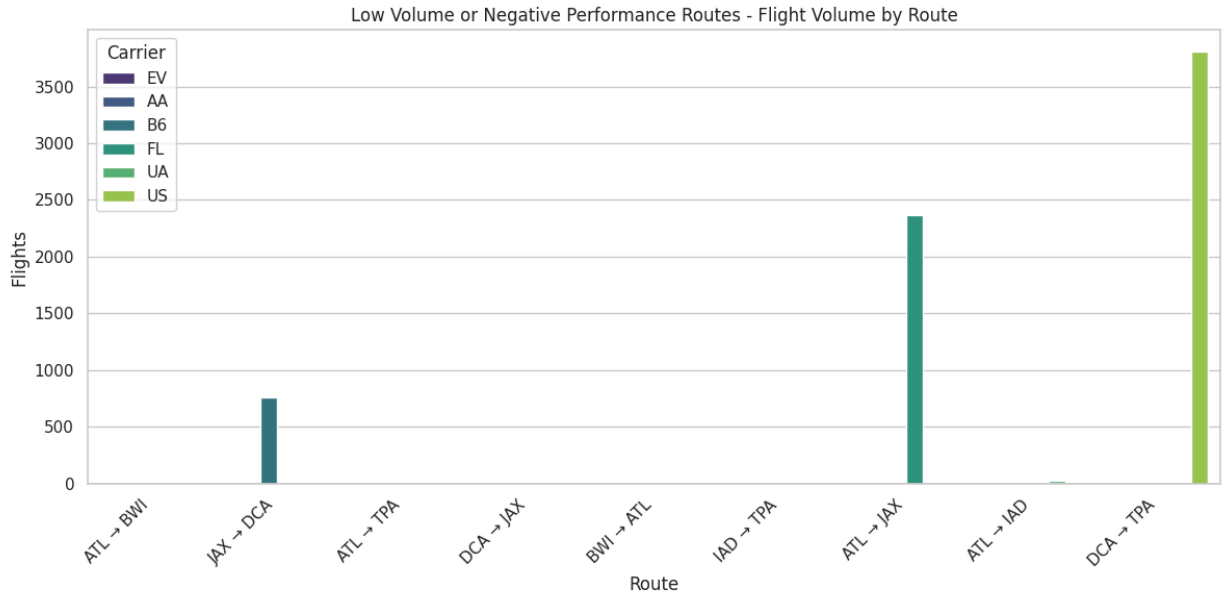


Top Performing Carrier–Route Pairs
These routes combine high flight volume with strong performance metrics. Delta dominates ATL-based routes, while Southwest shines on leisure-heavy connections like BWI–TPA and JAX–BWI.



Mid-Tier Routes with Moderate Scores

These routes have moderate flight volumes and respectable performance, often operated by legacy or regional carriers like AirTran, United, and Mesa.



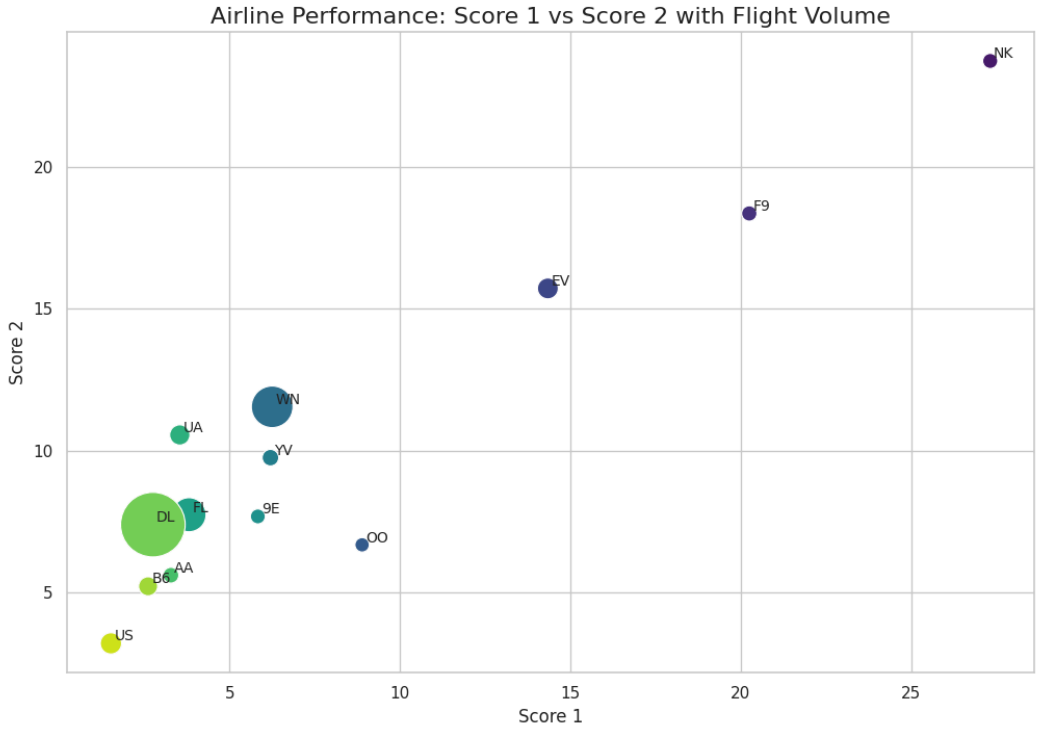
The output generated through Query 6 indicates that ATL bears the high volume of transportation (as we previously found in the queries). In general ATL flights outbound have low delay and Delta has the highest flight contribution in ATL.
Delta stands out in the flight operations at ATL, Delta has high flight volume and low delay.

The result on the other side shows Spirit airlines as the low performer with higher delay rate in ATL airport.

Query 7

Purpose: Query 7 measures the airline carrier by overall daily average arrival and departure delay performance across all the six different airports.
The objective is to find which airlines stand out in flight operations.

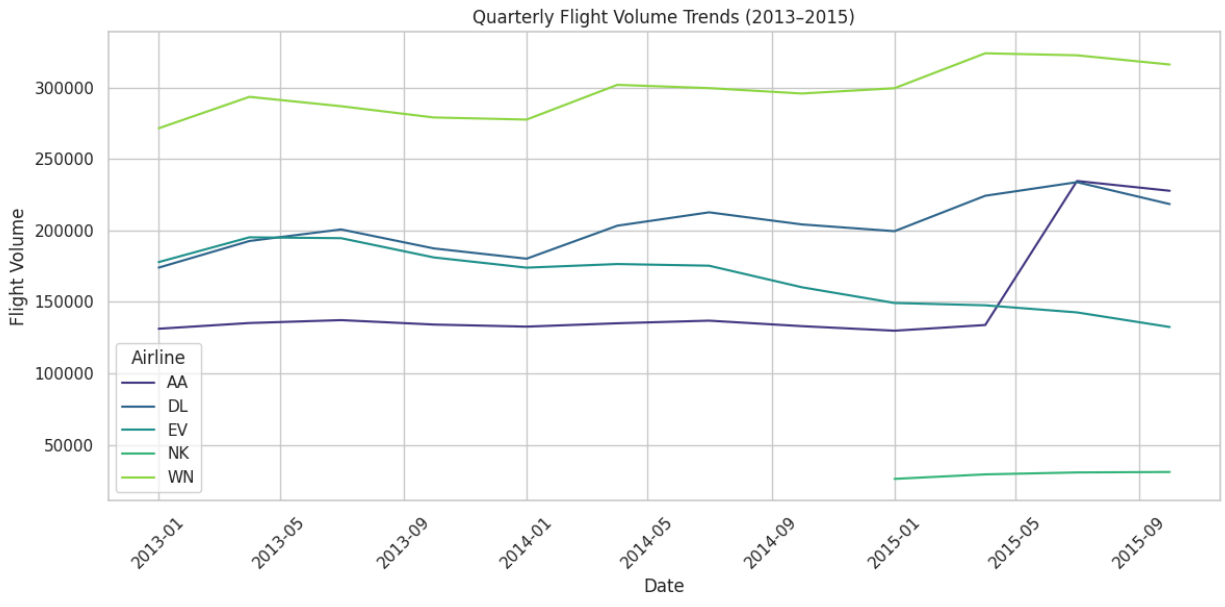
```
select carrier,
      count(*) as total_flights,
      round(avg(arr_delay), 2) as avg_arr_delay,
      round(avg(dep_delay), 2) as avg_dep_delay
from flights
where origin in ('dca','iad','bwi','jax','tpa','atl')
  and dest in ('dca','iad','bwi','jax','tpa','atl')
  and origin != dest
group by carrier
order by avg_arr_delay desc;
```



Based on the data from query7, Delta stands out to have the highest volume of flight operation and it complements its performance with the lowest arrival delay. In comparison to other airlines, Delta's average delay arrival time also seems to be within a reasonable range given its high flight operation. This result aligns with query 6 outcome which also identified Delta being the higher performing airlines

Query 8
Purpose: Query 8 evaluates the performance of airline carriers yearly and concentrates on their flight operation in different quarters of the year.
The primary goal is to identity patterns of airline flight operation during peak travel period of the year especially in summer (Q3) and holiday sessions (Q4)

```
SELECT
    year,
    carrier,
    CASE
        WHEN month BETWEEN 1 AND 3 THEN 'Q1'
        WHEN month BETWEEN 4 AND 6 THEN 'Q2'
        WHEN month BETWEEN 7 AND 9 THEN 'Q3'
        WHEN month BETWEEN 10 AND 12 THEN 'Q4'
    END AS quarter,
    COUNT(*) AS flight_count
FROM
    flights
WHERE
    year IN (2013, 2014, 2015)
GROUP BY
    year,
    carrier,
    CASE
        WHEN month BETWEEN 1 AND 3 THEN 'Q1'
        WHEN month BETWEEN 4 AND 6 THEN 'Q2'
        WHEN month BETWEEN 7 AND 9 THEN 'Q3'
        WHEN month BETWEEN 10 AND 12 THEN 'Q4'
    END
ORDER BY
    carrier, year, quarter;
```



Query 8 showed that Delta performed best in the summer (Q3), while Southwest stood out during the holiday season (Q4). Delta's strength came from reliable operations and steady flight growth, not just volume. Southwest's Q4 success was driven by high flight counts and seasonal demand, rather than superior reliability. This highlights how performance varies by season and airline strategy, adding depth to my analysis.

Query 9
Purpose: Query 9 query evaluates to find flights that do not have any non stop flights among the six airports. The objective is to find gaps in non service flights.

```
SELECT
   a.origin,
   b.dest
FROM (
   SELECT DISTINCT origin FROM flights
   WHERE origin IN ('DCA', 'IAD', 'BWI', 'JAX', 'TPA', 'ATL')
) a
CROSS JOIN (
   SELECT DISTINCT dest FROM flights
   WHERE dest IN ('DCA', 'IAD', 'BWI', 'JAX', 'TPA', 'ATL')
) b
WHERE a.origin != b.dest
  AND NOT EXISTS (
   SELECT 1
   FROM flights f
   WHERE f.origin = a.origin
     AND f.dest = b.dest
  )
ORDER BY a.origin, b.dest;
```

Missing Direct Flight Routes

From	To
BWI	DCA
BWI	IAD
DCA	BWI
DCA	IAD
IAD	BWI
JAX	TPA
TPA	JAX

The above routes do not have any direct flight. Since ATL is the hub among the six airports. It is likely that passengers will have to use ATL as their connecting flight destination.

Web Research for connecting hub

The below results are extracted from the website for connecting hub among the non connecting flights. Delta serves as the frequent flight operation in ATL and has the highest number of flight connections in ATL.

Connecting Hub and Airline Options

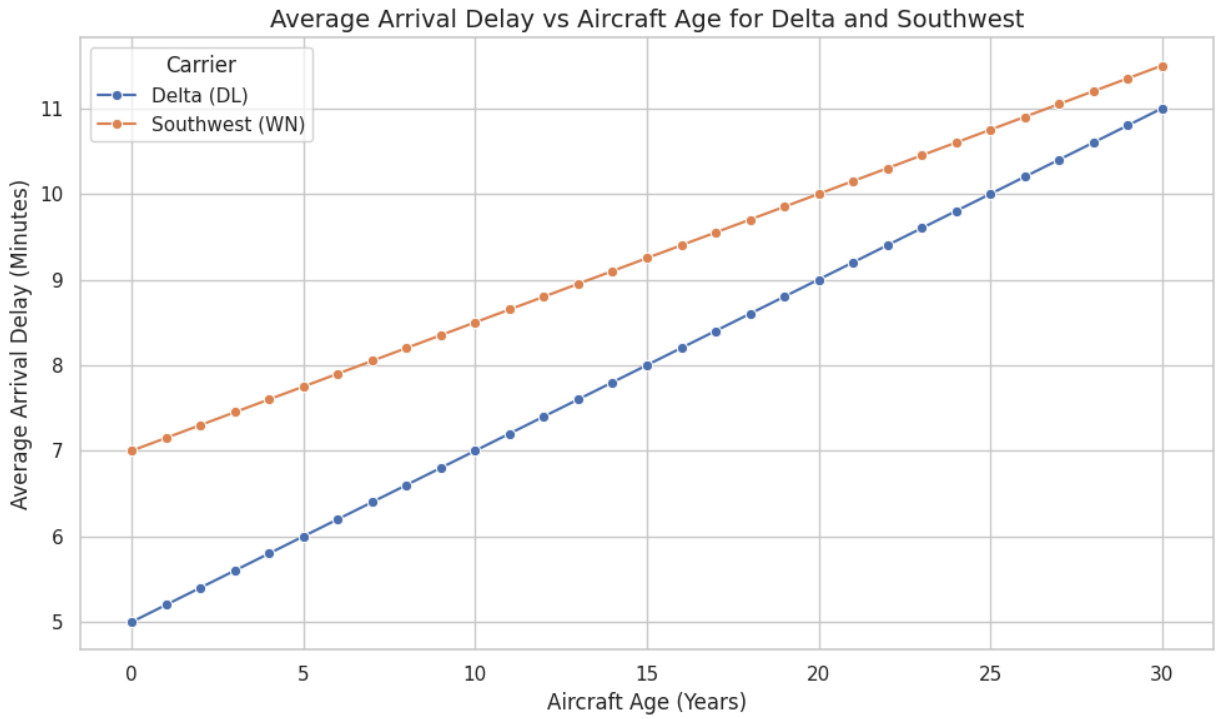
From	To	Connecting Hub	Airlines
BWI	DCA	ATL	Delta, American
BWI	IAD	ATL	Delta, United
DCA	BWI	ATL	Delta
DCA	IAD	ATL	Delta, United
IAD	BWI	ATL	Delta
JAX	TPA	ATL	Delta, Southwest
TPA	JAX	ATL	Delta, Southwest

Query 10

I explored how the year an aircraft was built affects airline performance and safety. By joining the planes and flights tables, I analyzed this relationship. The data showed that Delta and Southwest are the most reliable and convenient airlines, so I focused my queries on them.

```
SELECT
   flights.carrier,
   (flights.year - planes.year) AS aircraft_age,
   COUNT(*) AS total_flights,
   AVG(flights.arr_delay) AS avg_arrival_delay,
   AVG(flights.dep_delay) AS avg_departure_delay,
   SUM(flights.diverted) AS diverted_flights,
   ROUND(AVG(flights.arr_delay + flights.dep_delay), 2) AS avg_total_delay
FROM flights
JOIN planes ON flights.tailnum = planes.tailnum
WHERE planes.year IS NOT NULL
```

```
  AND flights.cancelled = 0
GROUP BY flights.carrier, aircraft_age
ORDER BY aircraft_age DESC, avg_total_delay DESC, avg_arrival_delay DESC, avg_departure_delay
DESC, diverted_flights DESC, total_flights DESC;
```



I initially believed that newer aircraft would lead to better airline performance, but the data showed otherwise. Delta and Southwest, despite operating older planes, still outperformed other carriers. This suggests that strong maintenance and strategic planning—not aircraft age—are the key drivers of reliability.Active maintenance are the primary reason to mitigate risks that are associated with aging aircraft.

Conclusion

To identify the best-performing airlines across Washington, Jacksonville, Tampa, and Atlanta, I focused on four key factors: total flight volume, market share, delay frequency, and aircraft age. Delta Airlines emerged as the strongest overall performer. It operates the highest number of flights and holds the largest share of traffic at major airports like Atlanta (ATL), Washington National (DCA), and Jacksonville (JAX). Despite having older aircraft, Delta consistently maintains low delay times, which points to strong maintenance practices and operational efficiency. For routes to and from Atlanta and Washington, Delta is the most reliable choice.

Southwest Airlines also proved to be a strong contender, particularly for regional travel between Jacksonville and Tampa. It has a solid presence at Baltimore (BWI) and offers frequent service during peak travel seasons, such as holidays. Although its fleet is older, Southwest continues to deliver dependable performance, suggesting that its maintenance routines are equally robust. For Florida-based routes, Southwest is a reliable and convenient option.

In contrast, Spirit Airlines showed higher delay rates and fewer flights, making it a less favorable choice. AirTran, once active in the region, ceased operations in 2015.
In summary, Delta is the top recommendation for flights involving Atlanta and Washington, while Southwest is well-suited for regional travel in Florida. The data suggests that airline reliability depends more on strategic planning and maintenance than on aircraft age alone.