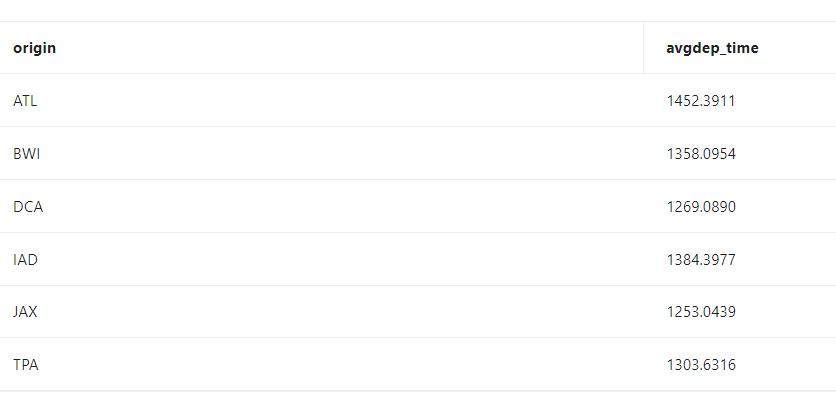
I. Initial Recommendation

1. Summary Review of Basis for the Recommendation (In General Terms)

Our analysis employed a multi-facetted approach considering the frequency of the flights, average flight duration, on time departure and arrival rates and incidents of flight disruptions. Data was sourced from the internal corporate records. We defined convenience metrics and performance metrics and I developed the formula to calculate performance score. The performance score is calculated by the ratio of on time flights, multiplied by the log of the total number of flights by each carrier. This way we consider not only the on time ratio, but also the total number of flights, which is an important factor. If we leave in a rather remote place maybe there is a local airline that has a good on time ratio, but the number of total flights may be much smaller compared to some major airlines. In those cases, we need to add the total number of flights as a factor in the performance score calculation. After the analysis and executing those queries, I found out that Delta Airlines has the best number of performance score, flying from and to those 4 cities (Jacksonville, Washington, Tampa, and Atlanta), so that is our recommendation for the corporate airline.

II.  Discussion of the Findings

1. According to the performance matrix based on convenience and performance and the frequency of the flight, we should choose the DL(Delta Airline) as our carrier for the corporate airline choice.



There is no dictionary for the dep\_time field so we do an average on that field to get a rough idea for the average departure time for those airports.

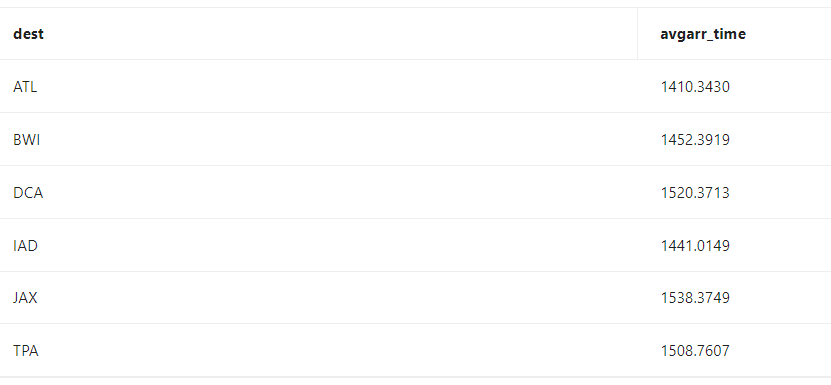
C.

select dest, avg(arr\_time) as avgarr\_time

from flights

where dest in ('DCA', 'IAD', 'BWI', 'JAX', 'TPA', 'ATL')

group by dest;



There is no dictionary for the arr\_time field so we do an average on that field to get a rough idea for the average arrival time for those airports.

D.

--distribution of group by hour of arrival time

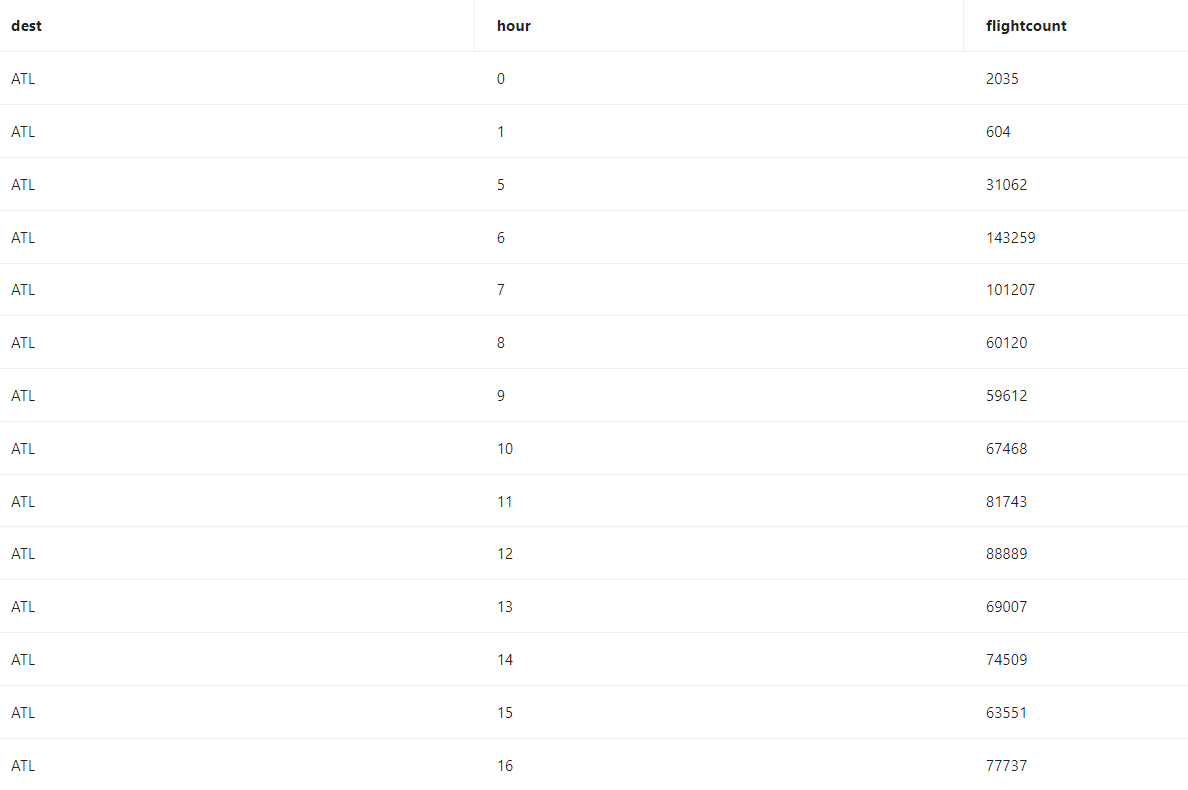
select dest, hour, count(\*) as flightcount

from flights

where dest in ('DCA', 'IAD', 'BWI', 'JAX', 'TPA', 'ATL')

group by dest, hour

order by dest, hour;



From all of the historical data, we find out the number of flight arrivals to a certain airport at a particular hour.

E. Departure

--distribution of group by hour of departure time

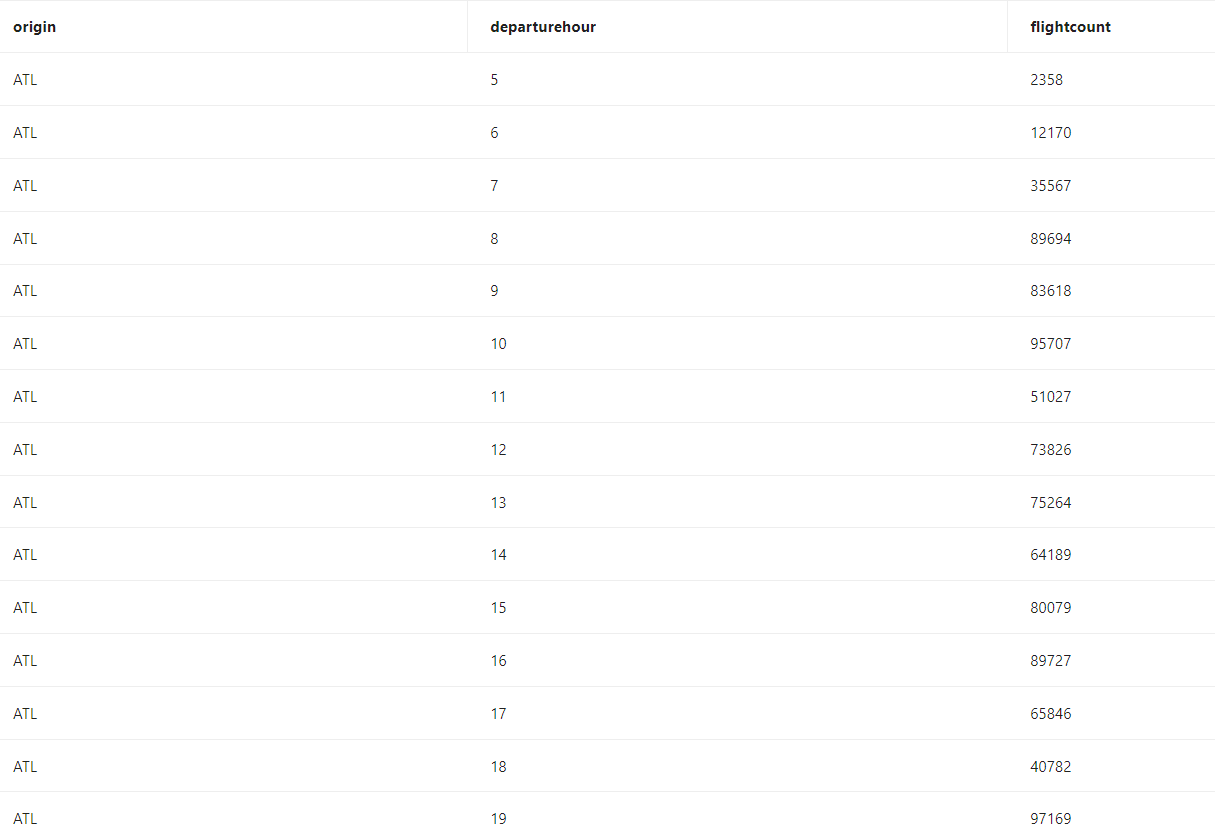
select origin, hour as departurehour, count(\*) as flightcount

from flights

where origin in ('DCA', 'IAD', 'BWI', 'JAX', 'TPA', 'ATL')

group by origin, departurehour

order by origin, departurehour;



From all of the historical data, we find out the number of flights departed from a certain airport at a particular hour.

F.



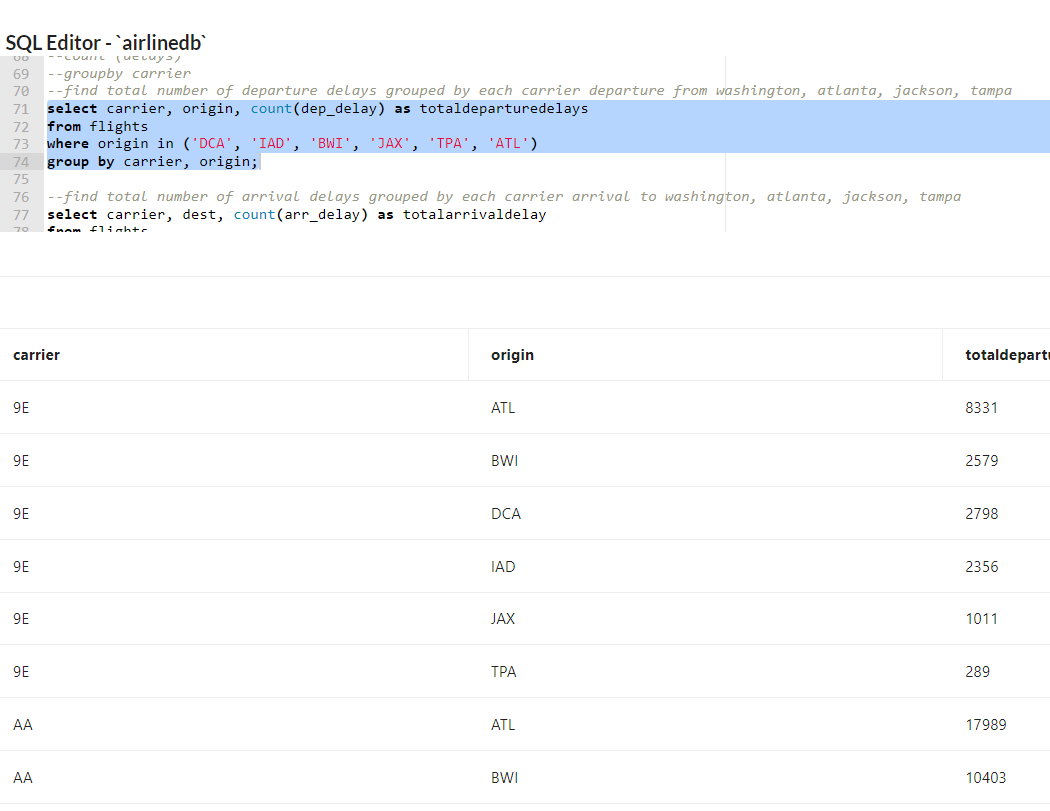
Find the average duration when departed from Washington DC, Tampa, Jacksonville, and Atlanta.

G.



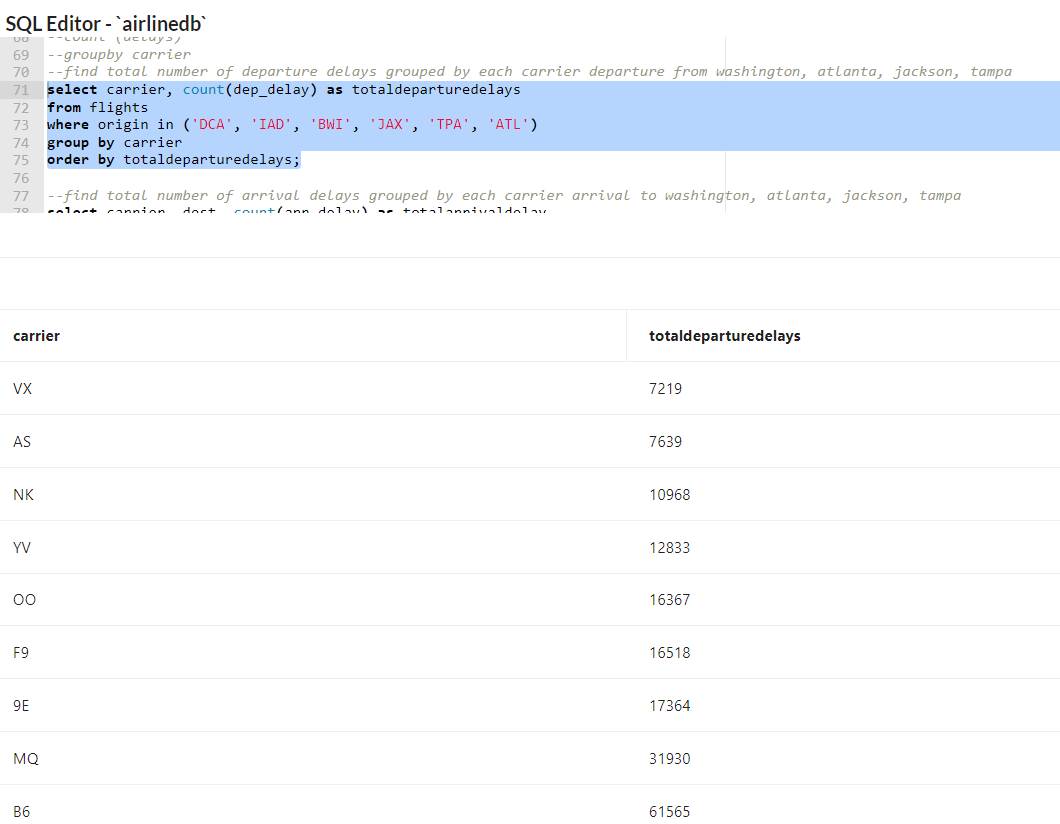
Find the average duration when arriving to Washington DC, Tampa, Jacksonville, and Atlanta.

H.



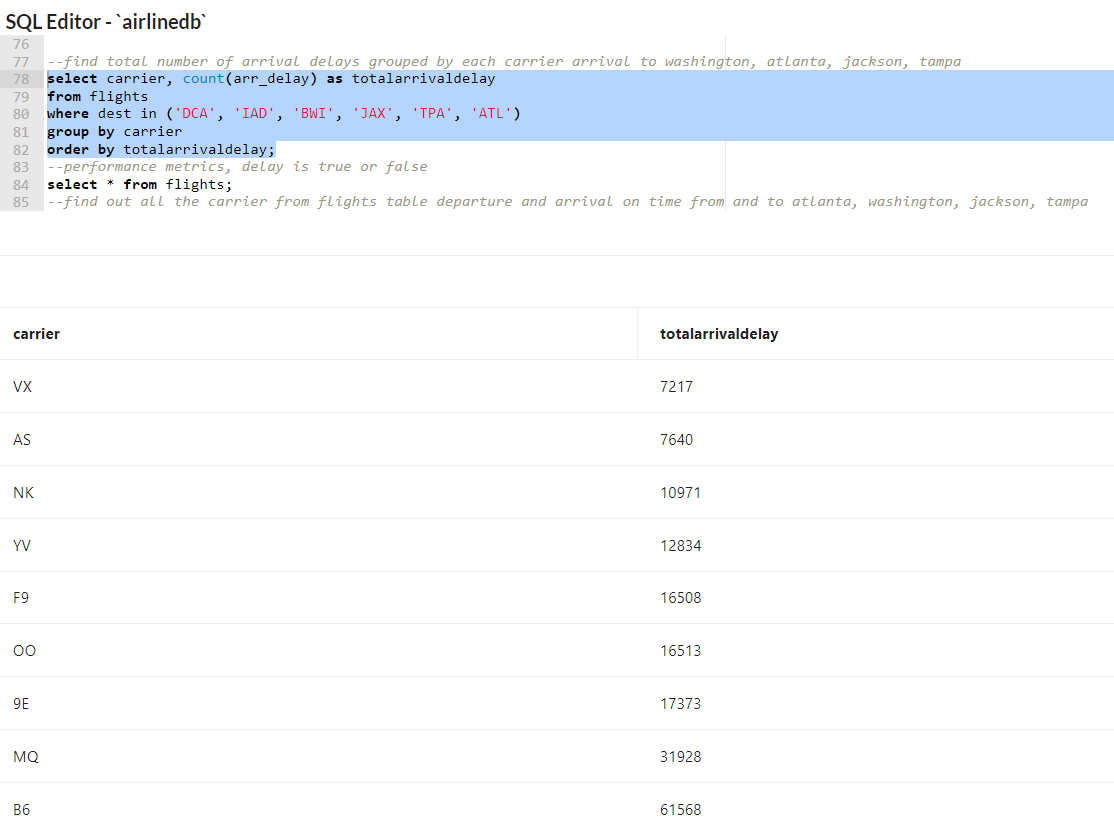
Find the total number of departure delays grouped by each carrier departure from Washington, Atlanta, Jacksonville, and Tampa.

I.



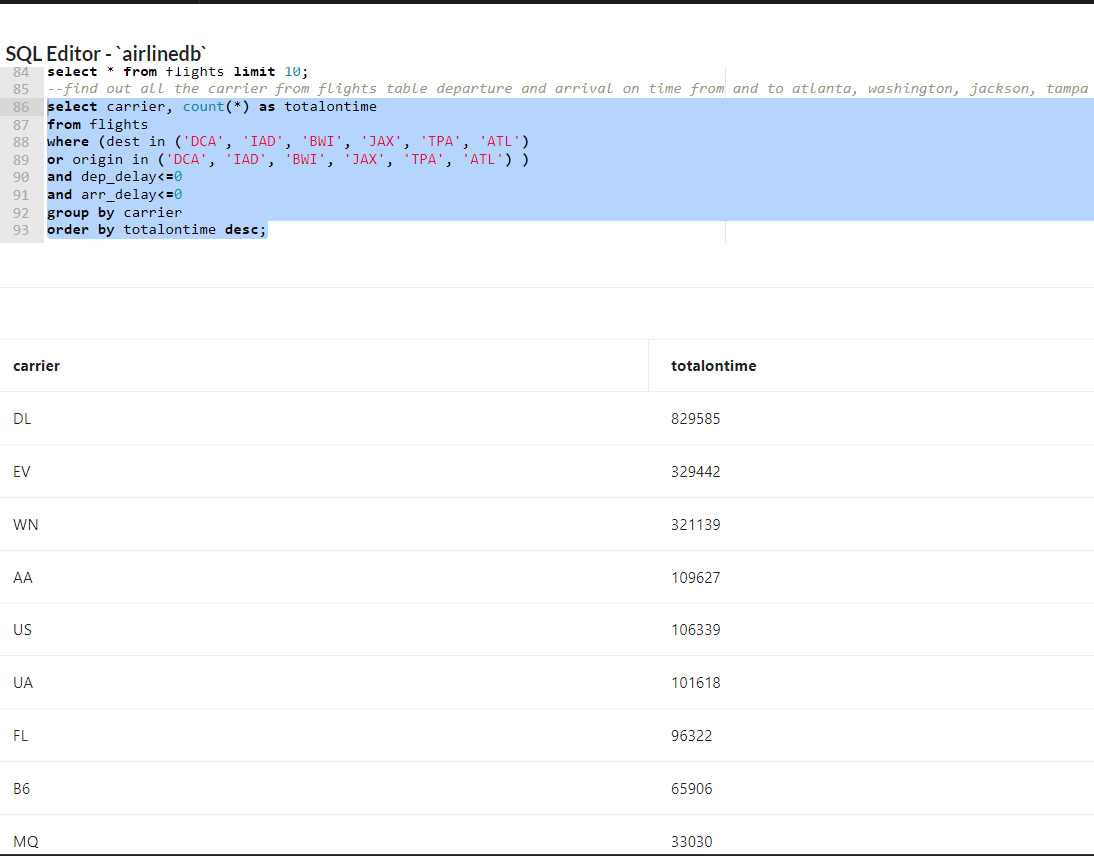
Find the total number of departure delays grouped by each carrier departure from Washington, Atlanta, Jacksonville, and Tampa.

J.



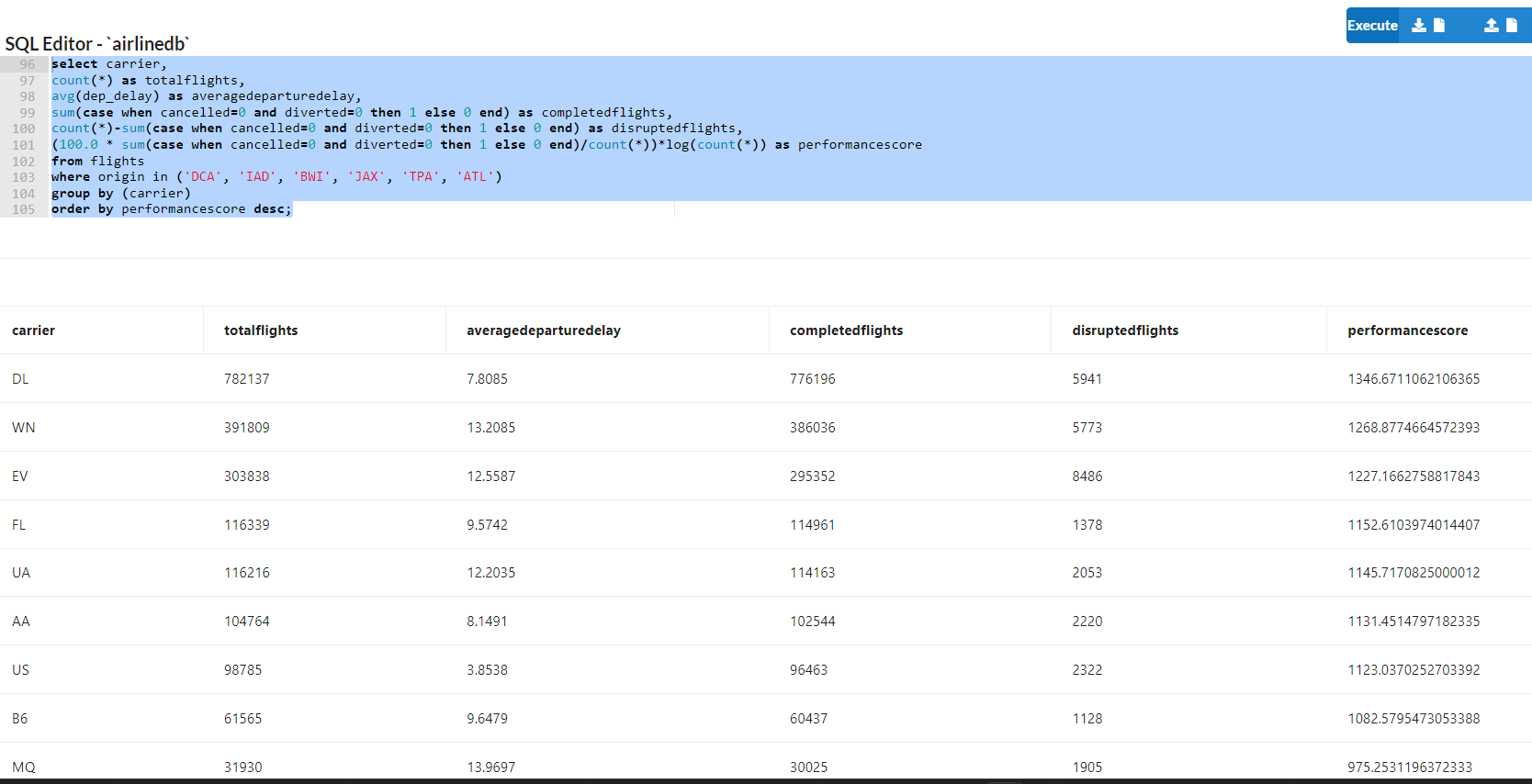
Find the total number of arrival delays grouped by each carrier arrival to Washington, Atlanta, Jacksonville, and Tampa.

K.



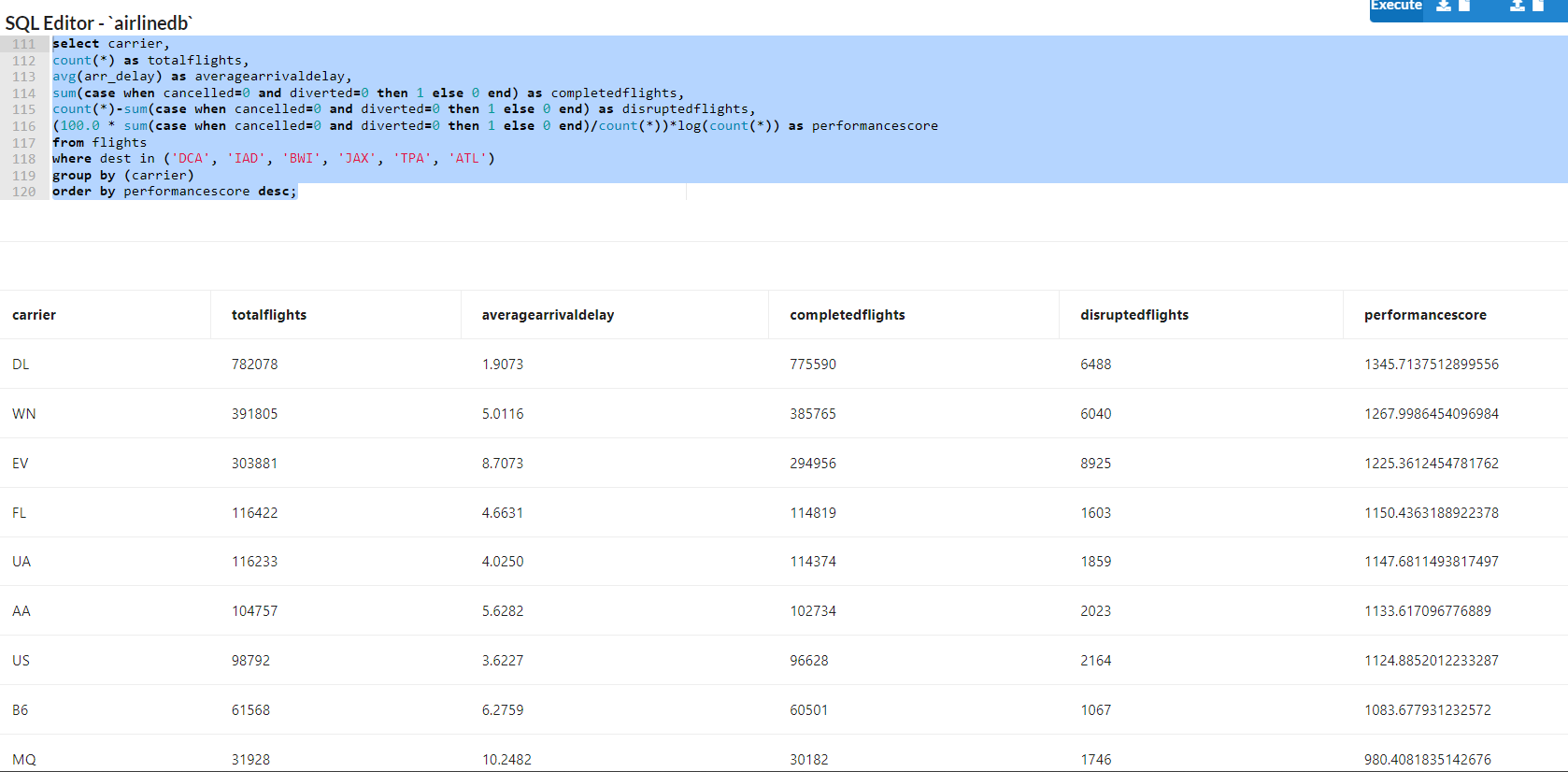
Find out all of the total number of flights for each carrier flight from Atlanta, Washington, Jacksonville, and Tampa that also arrived to Atlanta, Washington, Jacksonville, and Tampa on time.

L.



Find the best airline based on departure count, performance and convenience. The performance score is measured by the total number of on time flights divided by the total number of flights then multiply the log of the total number of flights.

M.



Find the best airline based on arrival count, performance and convenience. The performance score is measured by the total number of on time flights divided by the total number of flights then multiply the log of the total number of flights.

III.  Appendix Including Code (Annotated)

1. Might replicate or modify your analysis upon request

Describe flights;

select @@version;

--Select all the flights departure or arrived from Washington Dc, Tampa, Jacksonville, and Atlanta.

select flight, origin, dest, dep\_time, arr\_time, sched\_arr\_time,air\_time, carrier, cancelled, diverted

from flights

where origin in ('DCA', 'IAD', 'BWI', 'JAX', 'TPA', 'ATL')

or dest in ('DCA', 'IAD', 'BWI', 'JAX', 'TPA', 'ATL')

limit 10;

select \* from flights

limit 1000;

--Calculate convenience metrics, measure the frequency of the flights.

select dest as dailyflights

from flights

where origin in ('DCA', 'IAD', 'BWI', 'JAX', 'TPA', 'ATL');

or dest in ('DCA', 'IAD', 'BWI', 'JAX', 'TPA', 'ATL')

groupby dest;

--Original --find out the average department time for the flight departure from DC, TAMPA, ATLANTA, and JACKSONVILLE.

select origin, avg(dep\_time) as avgdep\_time

from flights

where origin in ('DCA', 'IAD', 'BWI', 'JAX', 'TPA', 'ATL')

group by origin;

--Looking for average departure time throughout the day.

select origin, sec\_to\_time(avg(time\_to\_sec(str\_to\_date(lpad(dep\_time, 4,'0'), '%H:%i') ))) as avgdep\_time

from flights

where origin in ('DCA', 'IAD', 'BWI', 'JAX', 'TPA', 'ATL')

group by origin;

--Looking for average arrival time throughout the day.

select dest, avg(arr\_time) as avgarr\_time

from flights

where dest in ('DCA', 'IAD', 'BWI', 'JAX', 'TPA', 'ATL')

group by dest;

--Distribution of group by hour of departure time.

select origin, hour as departurehour, count(\*) as flightcount

from flights

where origin in ('DCA', 'IAD', 'BWI', 'JAX', 'TPA', 'ATL')

group by origin, departurehour

order by origin, departurehour;

--Distribution of group by hour of arrival time.

select dest, hour, count(\*) as flightcount

from flights

where dest in ('DCA', 'IAD', 'BWI', 'JAX', 'TPA', 'ATL')

group by dest, hour

order by dest, hour;

--Check field name duration. Looking for average duration for flight departure from Washington, Tampa, Jacksonville, and Atlanta.

select origin, avg(air\_time) as averageduration

from flights

where origin in ('DCA', 'IAD', 'BWI', 'JAX', 'TPA', 'ATL')

group by origin;

describe flights;

--Check field name duration. Looking for average duration for flight arrival to Washington, Tampa, Jacksonville, and Atlanta.

select dest, avg(air\_time) as averageduration

from flights

where dest in ('DCA', 'IAD', 'BWI', 'JAX', 'TPA', 'ATL')

group by dest;

show tables;

--count (delays)

--groupby carrier

--Find the total number of departure delays grouped by each carrier departure from Washington, Atlanta, Jacksonville, and Tampa.

select carrier, count(dep\_delay) as totaldeparturedelays

from flights

where origin in ('DCA', 'IAD', 'BWI', 'JAX', 'TPA', 'ATL')

group by carrier

order by totaldeparturedelays;

--Find the total number of arrival delays grouped by each carrier arrival to Washington, Atlanta, Jacksonville, and Tampa.

select carrier, count(arr\_delay) as totalarrivaldelay

from flights

where dest in ('DCA', 'IAD', 'BWI', 'JAX', 'TPA', 'ATL')

group by carrier

order by totalarrivaldelay;

--Performance metrics, delay is true or false.

select \* from flights limit 10;

--Find out all the carrier from flights table departure and arrival on time from and to Atlanta, Washington, Jacksonville, Tampa.

select carrier, count(\*) as totalontime

from flights

where (dest in ('DCA', 'IAD', 'BWI', 'JAX', 'TPA', 'ATL')

or origin in ('DCA', 'IAD', 'BWI', 'JAX', 'TPA', 'ATL') )

and dep\_delay<=0

and arr\_delay<=0

group by carrier

order by totalontime desc;

--Cancel diverted, find the best airline based on departure count, performance and convenience.

select carrier,

count(\*) as totalflights,

avg(dep\_delay) as averagedeparturedelay,

sum(case when cancelled=0 and diverted=0 then 1 else 0 end) as completedflights,

count(\*)-sum(case when cancelled=0 and diverted=0 then 1 else 0 end) as disruptedflights,

(100.0 \* sum(case when cancelled=0 and diverted=0 then 1 else 0 end)/count(\*))\*log(count(\*)) as performancescore

from flights

where origin in ('DCA', 'IAD', 'BWI', 'JAX', 'TPA', 'ATL')

group by (carrier)

order by performancescore desc;

describe flights;

-- Find the best airline based on arrival cities, performance and convenience.

select carrier,

count(\*) as totalflights,

avg(arr\_delay) as averagearrivaldelay,

sum(case when cancelled=0 and diverted=0 then 1 else 0 end) as completedflights,

count(\*)-sum(case when cancelled=0 and diverted=0 then 1 else 0 end) as disruptedflights,

(100.0 \* sum(case when cancelled=0 and diverted=0 then 1 else 0 end)/count(\*))\*log(count(\*)) as performancescore

from flights

where dest in ('DCA', 'IAD', 'BWI', 'JAX', 'TPA', 'ATL')

group by (carrier)

order by performancescore desc;