

Laboratorio 3 SQL PARTE 2

Uso de join's

1. ¿Cuál es el estado del que salió el vuelo con el mayor retraso en el tiempo de salida registrado?
 - a. El aeropuerto de origen y el retraso en el tiempo de salida. De manera que muestra que el vuelo con un mayor retraso fue aquel con el código CLT y se atrasó 2467 minutos.

The screenshot shows the pgAdmin 4 interface. The left sidebar displays the database structure, including the 'public' schema and the 'flights' database. The main pane shows a SQL query:

```
SELECT origin, depdelay
FROM flight
WHERE depdelay is not null
ORDER BY depdelay DESC
LIMIT 1
```

The 'Data output' tab shows the results:

origin	depdelay
CLT	2467

The status bar at the bottom indicates: 'Total rows: 1 of 1', 'Query complete 00:00:03.396', and 'Successfully run. Total query runtime: 3 secs 396 msec. 1 rows affected.'

- b. Query que demuestra el estado, el código, el nombre del aeropuerto, y el país del aeropuerto.

The screenshot shows the pgAdmin 4 interface. The left sidebar displays the database structure, including the 'public' schema and the 'flights' database. The main pane shows a SQL query:

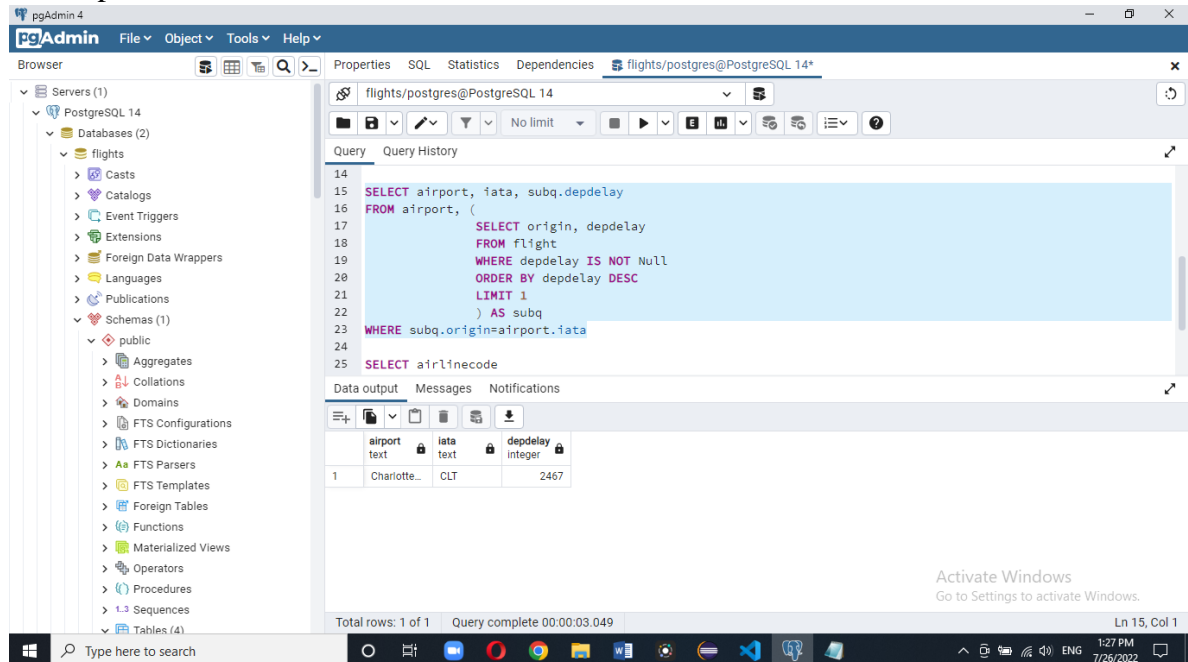
```
SELECT iata, airport, airport.state, airport.country
FROM airport, (
    SELECT origin, depdelay
    FROM flight
    WHERE depdelay IS NOT Null
    ORDER BY depdelay DESC
    LIMIT 1
) AS subq
WHERE subq.origin=airport.iata
SELECT airport, iata, subq.depdelay
```

The 'Data output' tab shows the results:

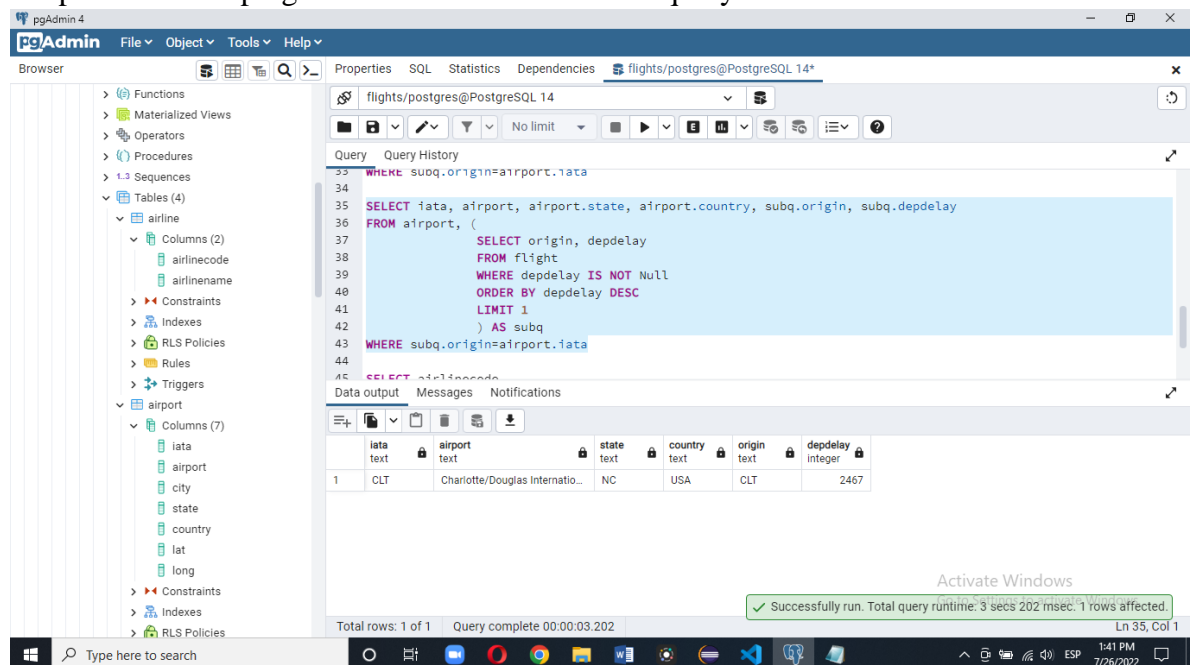
iata	airport	state	country
CLT	Charlotte...	NC	USA

The status bar at the bottom indicates: 'Total rows: 1 of 1', 'Query complete 00:00:03.270', and 'Go to Settings to activate Windows.'

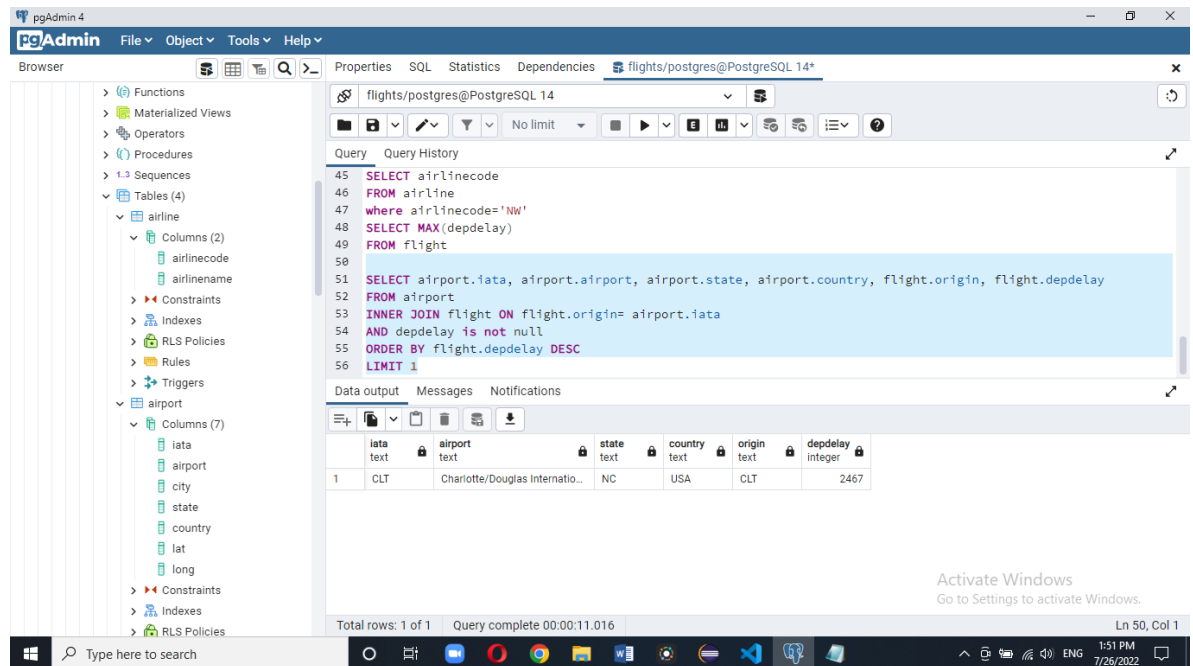
- c. Este sería una query que presenta el aeropuerto, el código del aeropuerto, y el tiempo de retraso.



2. Responda las dos preguntas anteriores con un solo query.



3. Responda finalmente la pregunta utilizando un solo query que haga uso de la sintaxis JOIN.



pgAdmin 4

flights/postgres@PostgreSQL 14*

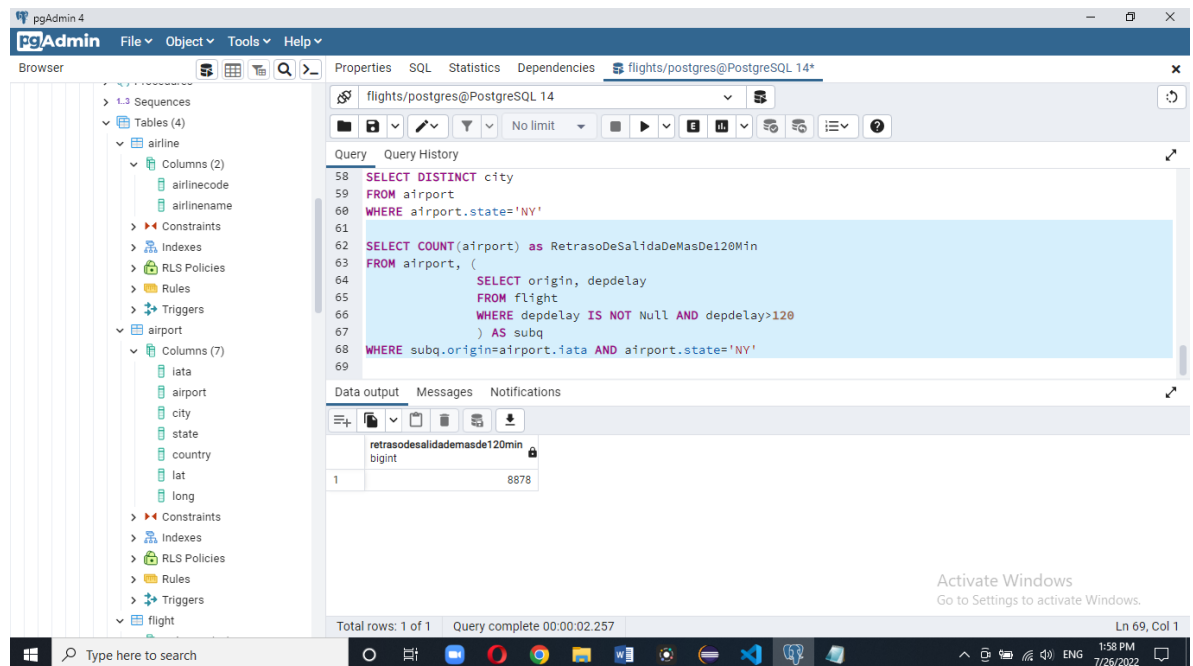
```
45 SELECT airlinecode
46 FROM airline
47 where airlinecode='NW'
48 SELECT MAX(depdelay)
49 FROM flight
50
51 SELECT airport.iata, airport.airport, airport.state, airport.country, flight.origin, flight.depdelay
52 FROM airport
53 INNER JOIN flight ON flight.origin= airport.iata
54 AND depdelay is not null
55 ORDER BY flight.depdelay DESC
56 LIMIT 1
```

iata	airport	state	country	origin	depdelay
CLT	Charlotte/Douglas Internatio...	NC	USA	CLT	2467

Total rows: 1 of 1 Query complete 00:00:11.016

Nótese que el tiempo de ejecución entre inner join y la subquery, es mayor para el inner join que la subquery.

4. ¿Cuántos vuelos tuvieron un retraso de salida de más de 120 minutos saliendo del estado de NY?



pgAdmin 4

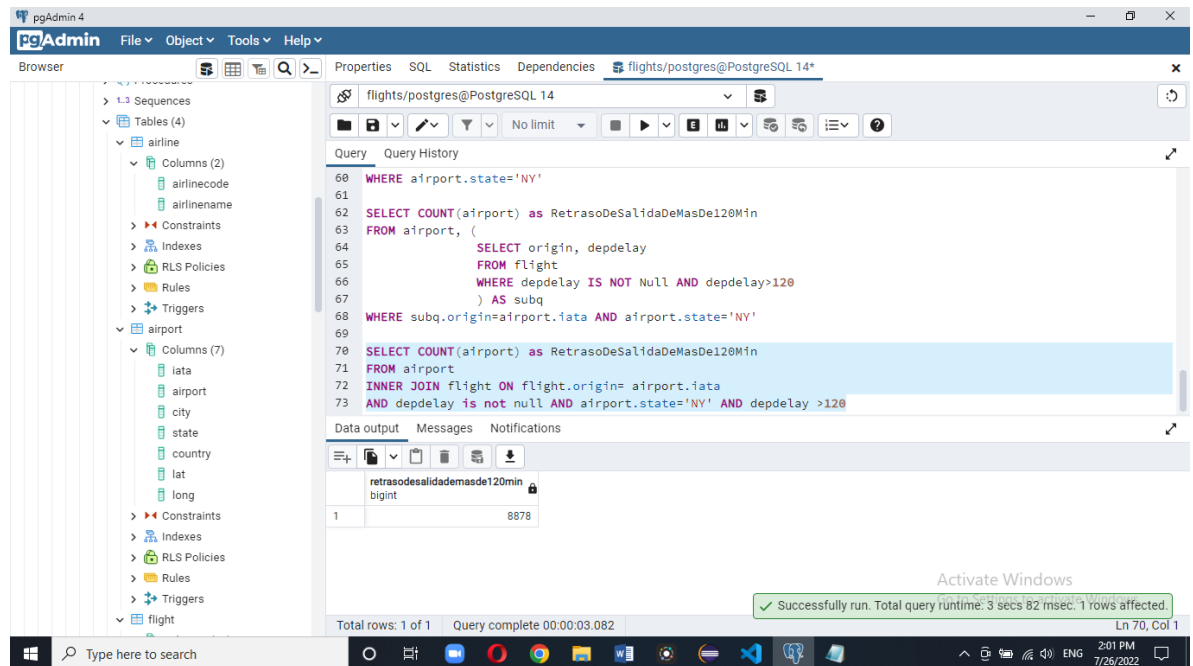
flights/postgres@PostgreSQL 14*

```
58 SELECT DISTINCT city
59 FROM airport
60 WHERE airport.state='NY'
61
62 SELECT COUNT(airport) as RetrasoDeSalidaDeMasDe120Min
63 FROM airport, (
64     SELECT origin, depdelay
65     FROM flight
66     WHERE depdelay IS NOT Null AND depdelay>120
67 ) AS subq
68 WHERE subq.origin=airport.iata AND airport.state='NY'
69
```

retrasosalidademasde120min
8878

Total rows: 1 of 1 Query complete 00:00:02.257

Fueron un total de 8878 vuelos que tuvieron un retraso saliendo desde NY. Usando subquery.



The screenshot shows the pgAdmin 4 interface. On the left, the 'Browser' pane displays the database structure: Servers (1) > PostgreSQL 14 > Databases (2) > flights > Tables (4) > airport. The main pane shows a SQL query with line numbers 60 to 73. The query filters for flights from New York (NY) with a delay greater than 120 minutes, using an inner join between the 'flight' and 'airport' tables. The results pane shows a single row with the value 8878 for the 'retrasodesalidadesmasde120min' column.

```
60 WHERE airport.state='NY'
61
62 SELECT COUNT(airport) as RetrasoDeSalidaDeMasDe120Min
63 FROM airport, (
64     SELECT origin, depdelay
65     FROM flight
66     WHERE depdelay IS NOT Null AND depdelay>120
67 ) AS subq
68 WHERE subq.origin=airport.iata AND airport.state='NY'
69
70 SELECT COUNT(airport) as RetrasoDeSalidaDeMasDe120Min
71 FROM airport
72 INNER JOIN flight ON flight.origin= airport.iata
73 AND depdelay is not null AND airport.state='NY' AND depdelay >120
```

retrasodesalidadesmasde120min
8878

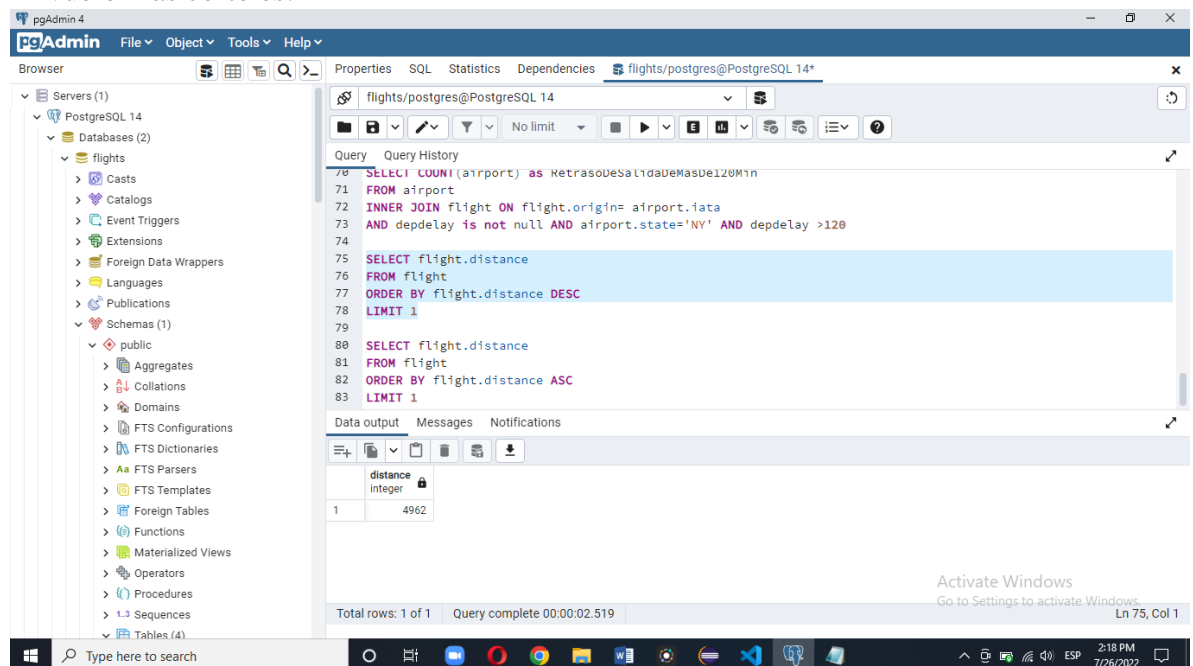
Successfully run. Total query runtime: 3 secs 82 msec. 1 rows affected.

Son 8878 vuelos con retraso, partiendo desde NY, utilizando el inner join.

Queries, Join's y agregaciones:

1. ¿Cuál es el vuelo más largos y más cortos en distancia registrados en la base de datos?

El vuelo más corto es:



The screenshot shows the pgAdmin 4 interface. On the left, the 'Browser' pane displays the database structure: Servers (1) > PostgreSQL 14 > Databases (2) > flights > Tables (4) > flight. The main pane shows a SQL query with line numbers 70 to 83. The query finds the shortest and longest flights by distance using inner joins and order by clauses. The results pane shows a single row with the value 4962 for the 'distance' column.

```
70 SELECT COUNT(airport) as RetrasoDeSalidaDeMasDe120Min
71 FROM airport
72 INNER JOIN flight ON flight.origin= airport.iata
73 AND depdelay is not null AND airport.state='NY' AND depdelay >120
74
75 SELECT flight.distance
76 FROM flight
77 ORDER BY flight.distance DESC
78 LIMIT 1
79
80 SELECT flight.distance
81 FROM flight
82 ORDER BY flight.distance ASC
83 LIMIT 1
```

distance
4962

Successfully run. Total query runtime: 00:00:02.519

El vuelo más largo es:

The screenshot shows the pgAdmin 4 interface. The left sidebar displays the database structure for 'PostgreSQL 14', including 'Databases (2)' and 'Schemas (1)'. The main pane shows a SQL query in the 'Query' tab:

```
70 SELECT COUNT(airport) as RetrasosDeSalidaDeMasDe120Mn
71 FROM airport
72 INNER JOIN flight ON flight.origin= airport.iata
73 AND depdelay is not null AND airport.state='NY' AND depdelay >120
74
75 SELECT flight.distance
76 FROM flight
77 ORDER BY flight.distance DESC
78 LIMIT 1
79
80 SELECT flight.distance
81 FROM flight
82 ORDER BY flight.distance ASC
83 LIMIT 1
```

The 'Data output' pane shows the results of the query:

distance	integer
1	11

The status bar at the bottom indicates 'Total rows: 1 of 1' and 'Query complete 00:00:03.423'.

2. ¿Qué aeropuertos no han tenido vuelos de salida?

The screenshot shows the pgAdmin 4 interface. The left sidebar displays the database structure for 'PostgreSQL 14', including 'Databases (2)' and 'Schemas (1)'. The main pane shows a SQL query in the 'Query' tab:

```
127 LIMIT 1
128
129 SELECT airport.iata, airport.airport, airport.state
130 FROM airport, flight
131 WHERE airport.iata NOT IN (flight.origin)
132
133 SELECT airport.iata, airport.airport, airport.state
134 FROM airport
135 FULL JOIN flight ON flight.origin = airport.iata
136 WHERE flight.origin IS null
137
138
139 SELECT DISTINCT airport.iata
```

The 'Data output' pane shows the results of the query:

iata	airport	state
1 0.00E+00	Moriarty	NM
2 0.00E+00	Crownpoint	NM
3 00M	Thigpen	MS
4 00R	Livingston Municipal	TX
5 00V	Meadow Lake	CO
6 01G	Perry-Warsaw	NY

The status bar at the bottom indicates 'Total rows: 1000 of 3007' and 'Query complete 00:00:01.808'. A green message box at the bottom right states: 'Successfully run. Total query runtime: 1 secs 808 msec: 3007 rows affected.'

Por ende hay 3007 aeropuertos sin vuelos de salida.

3. ¿Cuál es la temporada alta? De manera que Julio y Marzo es la temporada alta.

pgAdmin 4

flights/postgres@PostgreSQL 14*

```

Query
86 FROM airport
87
88 SELECT airport.
89
90 SELECT airport.iata, airport.airport, airport.state
91 FROM airport
92
93 SELECT flight.month, COUNT(flight.month)
94 FROM flight
95 GROUP BY flight.month
96 ORDER BY COUNT(flight.month) DESC
97 LIMIT 2
98
99

```

month	integer	count	bigint
1	7	62793	
2	3	61609	

Total rows: 2 of 2 Query complete 00:00:01.807

Successfully run. Total query runtime: 1 secs 807 msec. 2 rows affected.

4. ¿Cuáles son los peores días para viajar? De manera que el peor día para viajar es viernes, dado que se retrasa 10.95 días.

pgAdmin 4

flights/postgres@PostgreSQL 14*

```

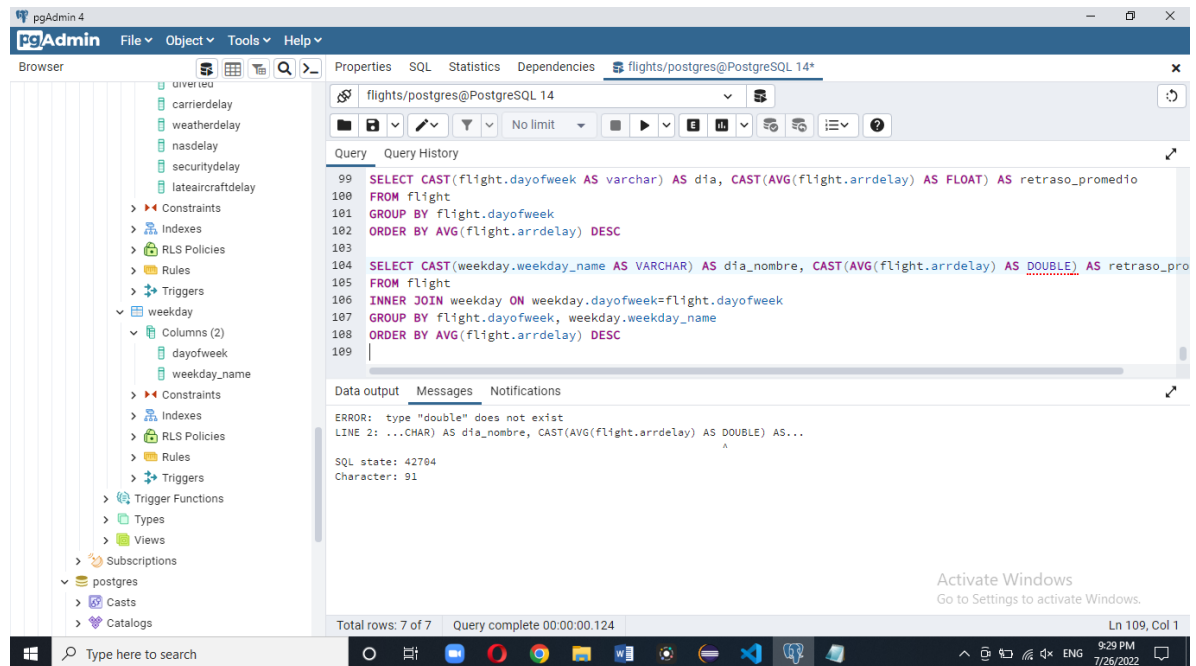
98
99 SELECT CAST(flight.dayofweek AS varchar) AS dia, CAST(AVG(flight.arrdelay) AS FLOAT) AS retraso_promedio
100 FROM flight
101 GROUP BY flight.dayofweek
102 ORDER BY AVG(flight.arrdelay) DESC
103
104 SELECT weekday.weekday_name AS dia, AVG(flight.arrdelay) AS retraso_promedio
105 FROM flight
106 INNER JOIN weekday ON weekday.dayofweek=flight.dayofweek
107 GROUP BY flight.dayofweek, weekday.weekday_name
108 ORDER BY AVG(flight.arrdelay) DESC
109

```

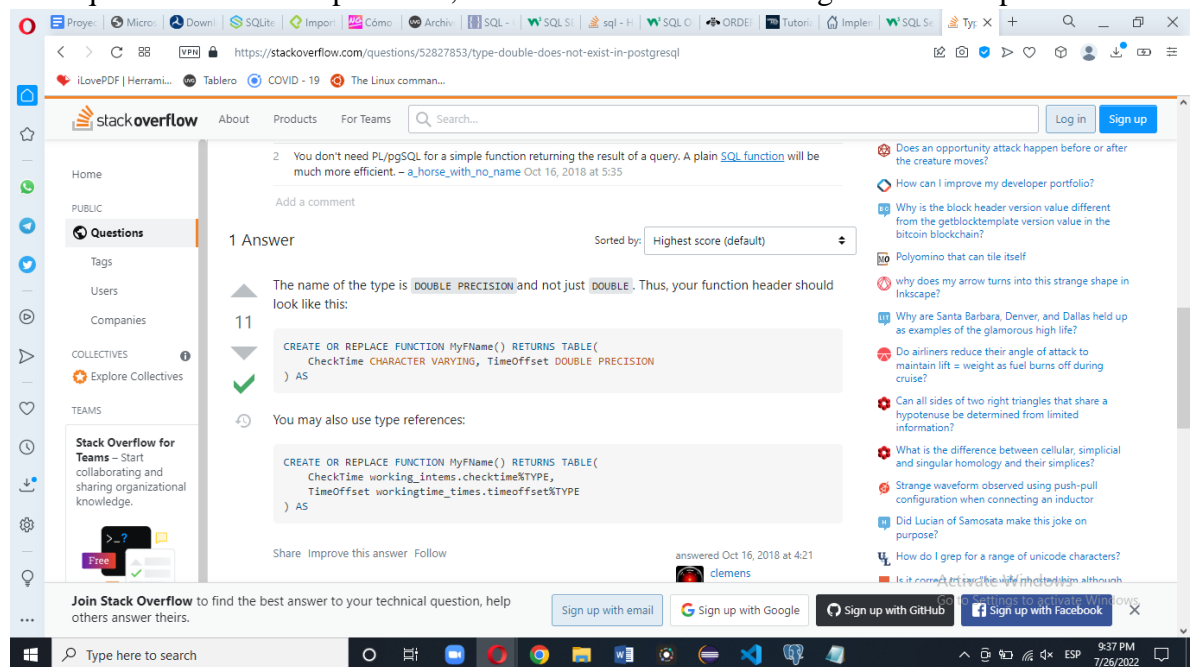
dia	character varying	retraso_promedio	double precision
1	5	10.953440079613667	
2	7	9.495886006775251	
3	4	8.41159915807781	
4	1	8.210850494863875	
5	2	7.481207604036612	
6	3	6.522017315719349	
7	6	5.789666410427927	

Total rows: 7 of 7 Query complete 00:00:03.241

5. ¿Cuáles son los peores días para viajar?



Como se observa da error debido a que no hay vartype de double en postgres sql, se tiene que colocar double precision, tal como se observa en la siguiente fotocaptura.



Por ende:

The screenshot shows the pgAdmin 4 interface. The left sidebar displays the database structure, including the 'weekday' table. The main pane shows a SQL query that calculates the average arrival delay for each day of the week. The query is as follows:

```

SELECT CAST(flight.dayofweek AS varchar) AS dia, CAST(AVG(flight.arrdelay) AS FLOAT) AS retraso_promedio
FROM flight
GROUP BY flight.dayofweek
ORDER BY AVG(flight.arrdelay) DESC

SELECT CAST(weekday.weekday_name AS VARCHAR) AS dia_nombre, CAST(AVG(flight.arrdelay) AS DOUBLE PRECISION) AS r
FROM flight
INNER JOIN weekday ON weekday.dayofweek=flight.dayofweek
GROUP BY flight.dayofweek, weekday.weekday_name
ORDER BY AVG(flight.arrdelay) DESC

```

The 'Data output' pane shows the results of the query:

	dia_nombre	retraso_promedio
1	Friday	10.9534400796136
2	Sunday	9.4958600677525
3	Thursday	8.41159915807781
4	Monday	8.21085049486387
5	Tuesday	7.48120760403661
6	Wednesday	6.52201731571934
7	Saturday	5.78966641042792

The status bar indicates: 'Successfully run. Total query runtime: 4 secs 421 msec. 7 rows affected. Ln 104, Col 1'.

Por ende, el viernes es el día con peor retraso promedio.

6. ¿Cuál es el vuelo más común?

The screenshot shows the pgAdmin 4 interface. The left sidebar displays the database structure, including the 'flight' table. The main pane shows a SQL query that counts the number of flights for each origin and destination. The query is as follows:

```

SELECT flight.origin, flight.dest, COUNT(flight.origin) AS CantidadDeVuelosCuyoOrigenYDestinoEsIgual
FROM flight
GROUP BY flight.origin, flight.dest
ORDER BY COUNT(flight.flightnum) DESC

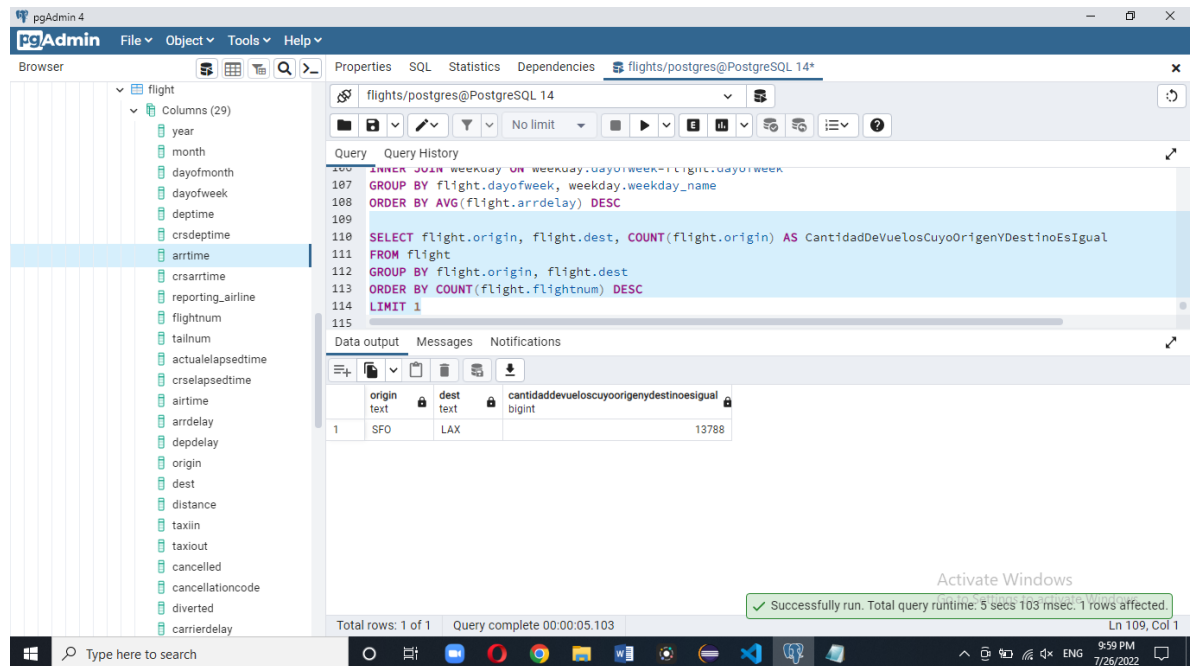
```

The 'Data output' pane shows the results of the query:

	origin	dest	cantidaddevueloscuyoorigenydestinoesigual
2	LAX	SFO	13390
3	OGG	HNL	12383
4	LGA	BOS	12035
5	BOS	LGA	12029
6	HNL	OGG	12014
7	LAX	LAS	11773
8	LAS	LAX	11729
9	LAX	SAN	11257
10	SAN	LAX	11224

The status bar indicates: 'Total rows: 1000 of 5366 Query complete 00:00:05.380 Ln 110, Col 2'.

Por ende el vuelo más común es:



The screenshot shows the pgAdmin 4 interface. On the left, the 'flight' table is selected in the 'Columns (29)' list. The main pane displays a SQL query and its results. The query is:

```
106 INNER JOIN weekday ON weekday.dayofweek=flight.dayofweek
107 GROUP BY flight.dayofweek, weekday.weekday_name
108 ORDER BY AVG(flight.arrdelay) DESC
109
110 SELECT flight.origin, flight.dest, COUNT(flight.origin) AS CantidadDeVuelosCuyoOrigenYDestinoEsIgual
111 FROM flight
112 GROUP BY flight.origin, flight.dest
113 ORDER BY COUNT(flight.flightnum) DESC
114 LIMIT 1
115
```

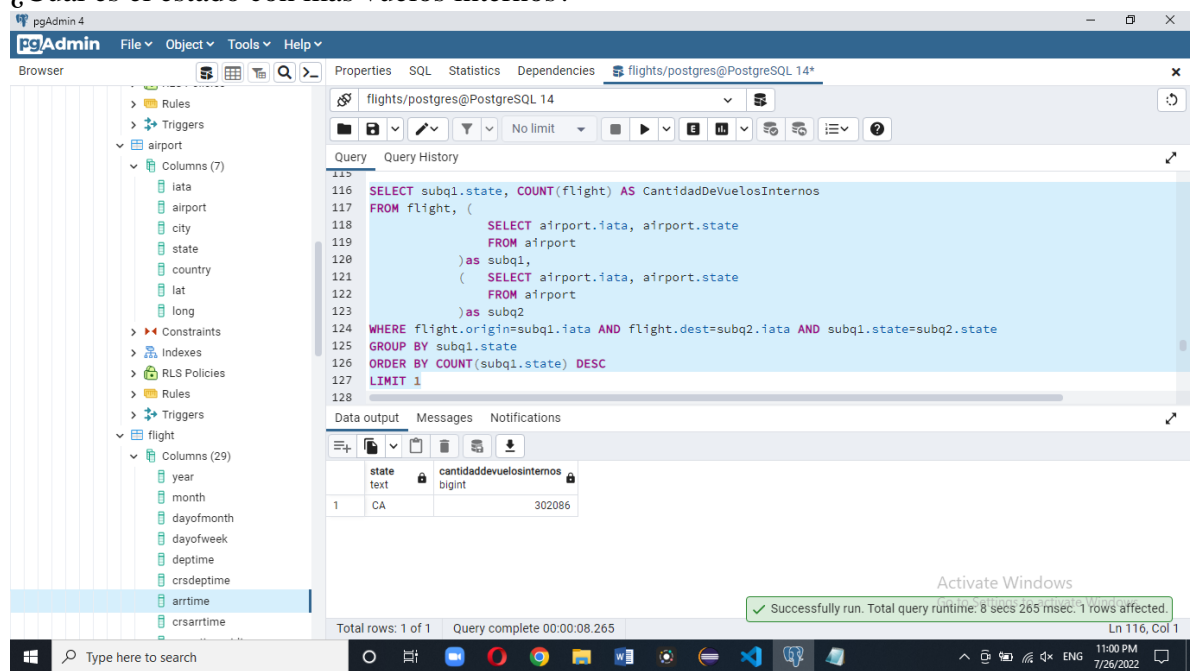
The 'Data output' tab shows the following result:

origin	dest	cantidaddevueloscuyoorigenydestinoesigual
SFO	LAX	13788

The status bar indicates: 'Total rows: 1 of 1', 'Query complete 00:00:05.103', and 'Successfully run. Total query runtime: 5 secs 103 msec. 1 rows affected.'

El vuelo de SFO a LAX con 13788 vuelos repetidos.

7. ¿Cuál es el estado con más vuelos internos?



The screenshot shows the pgAdmin 4 interface. On the left, the 'flight' table is selected in the 'Columns (29)' list. The main pane displays a SQL query and its results. The query is:

```
115
116 SELECT subq1.state, COUNT(flight) AS CantidadDeVuelosInternos
117 FROM flight, (
118     SELECT airport.iata, airport.state
119     FROM airport
120     )as subq1,
121     ( SELECT airport.iata, airport.state
122     FROM airport
123     )as subq2
124 WHERE flight.origin=subq1.iata AND flight.dest=subq2.iata AND subq1.state=subq2.state
125 GROUP BY subq1.state
126 ORDER BY COUNT(subq1.state) DESC
127 LIMIT 1
128
```

The 'Data output' tab shows the following result:

state	cantidaddevuelosinternos
CA	302086

The status bar indicates: 'Total rows: 1 of 1', 'Query complete 00:00:08.265', and 'Successfully run. Total query runtime: 8 secs 265 msec. 1 rows affected.'

El estado con más vuelos internos es CA con 302086 vuelos internos.

8. ¿Qué aerolíneas se retrasan más de 10 minutos regularmente?

The screenshot shows the pgAdmin 4 interface. The left pane displays the 'Columns (29)' list for the 'flights' table. The central pane shows a SQL query with line numbers 145 to 157. The query filters for airlines with an average arrival delay of 10 minutes or more. The bottom pane shows the 'Data output' tab with a table of results.

```
145 WHERE airline.airlinecode = flight.reporting_airline
146 GROUP BY airline.airlinename
147 HAVING AVG(flight.arrdelay) >= 10
148 ORDER BY AVG(flight.arrdelay) DESC
149 LIMIT 10
150
151 SELECT CAST(airline.airlinename AS VARCHAR) AS airline_name, CAST(AVG(flight.arrdelay) AS FLOAT) AS average_delay
152 FROM flight, airline
153 WHERE airline.airlinecode = flight.reporting_airline
154 GROUP BY airline.airlinename
155 HAVING AVG(flight.arrdelay) >= 10
156 ORDER BY AVG(flight.arrdelay) DESC
157 LIMIT 10
```

	airline_name character varying	average_delay double precision
1	American Airlines Inc.	12.607194035713981
2	PSA Airlines Inc.	11.817467683998007
3	Mesa Airlines Inc.	11.77518143360081
4	United Air Lines Inc.	11.291322186680183
5	JetBlue Airways	11.08418439051813
6	Continental Air Lines Inc.	10.9790372802913

Total rows: 8 of 8 Query complete 00:00:05.721

De manera que las aerolíneas que se retrasan en la llegada más de 10 minutos normalmente son:

- American Airlines Inc.
- PSA Airlines Inc.
- Mesa Airlines Inc.
- United Air Lines Inc.
- JetBlue Airways
- Continental Air Lines Inc.
- ExpressJet Airlines Inc.
- ExpressJet Airlines Inc. (1)

i. ExpressJet Airlines Inc.

The screenshot shows the pgAdmin 4 interface. On the left, a tree view lists columns for a table, including 'year', 'month', 'dayofmonth', 'dayofweek', 'deptime', 'crsdeptime', 'arrtime', 'crsarrtime', 'reporting_airline', 'flightnum', 'tailnum', 'actualelapsedtime', 'crselapsedtime', 'airtime', 'arrdelay', 'depdelay', 'origin', 'dest', 'distance', 'taxin', 'taxiout', 'cancelled', 'cancellationcode', 'diverted', 'carrierdelay', and 'weatherdelay'. The main pane displays a SQL query and its results.

Query:

```

140 FROM flight
141 LIMIT 10
142
143 SELECT CAST(airline.airlinename AS VARCHAR) AS airline_name, CAST(AVG(flight.depdelay) AS FLOAT) AS average_del
144 FROM flight, airline
145 WHERE airline.airlinecode = flight.reporting_airline
146 GROUP BY airline.airlinename
147 HAVING AVG(flight.depdelay) >= 10
148 ORDER BY AVG(flight.depdelay) DESC
149 LIMIT 10
150
151 SELECT CAST(airline.airlinename AS VARCHAR) AS airline_name, CAST(AVG(flight.arrdelay) AS FLOAT) AS average_del
152 FROM flight, airline
153 WHERE airline.airlinecode = flight.reporting_airline
154 GROUP BY airline.airlinename
155 HAVING AVG(flight.arrdelay) >= 10
156 ORDER BY AVG(flight.arrdelay) DESC
157 LIMIT 10

```

Data output:

	airline_name	average_delay
1	United Air Lines Inc.	14.11257661236138
2	American Airlines Inc.	13.280898264437912
3	Continental Air Lines Inc.	13.18522978602152
4	JetBlue Airways	12.653395748122113
5	Mesa Airlines Inc.	12.000675279875033
6	ExpressJet Airlines Inc.	11.922537970871462

At the bottom, a status bar indicates: "Total rows: 10 of 10 Query complete 00:00:06.194". A green message box says: "Successfully run. Total query runtime: 6 secs 194 msec: 10 rows affected. Ln 142, Col 1".

Mientras que las aerolíneas que se retrasan en el tiempo de despegue son:

- j. United Air Lines Inc.
- k. American Airlines Inc.
- l. Continental Air Lines Inc.
- m. JetBlue Airways
- n. Mesa Airlines Inc.
- o. ExpressJet Airlines Inc.
- p. PSA Airlines Inc.
- q. ExpressJet Airlines Inc. (1)