

## Peer-Graded Assignment: Analyzing Big Data with SQL

**Name:** Hoang Trung Nghia

**Date:** 29/09/2021

*(Include your name and today's date above.)*

### Assignment

Recommend which pair of United States airports should be connected with a high-speed passenger rail tunnel. To do this, write and run a SELECT statement to return pairs of airports that are between **300** and **400** miles apart and that had at least **5,000** (five thousand) flights per year on average *in each direction* between them. Arrange the rows to identify which one of these pairs of airports has largest total number of seats on the planes that flew between them. Your SELECT statement must return all the information required to fill in the table below.

### Recommendation

I recommend the following tunnel route:

	First Direction	Second Direction
Three-letter airport code for origin	SFO	LAX
Three-letter airport code for destination	LAX	SFO
Average flight distance in miles	337	337
Average number of flights per year	14712	14504
Average annual passenger capacity	1996597	1981059
Average arrival delay in minutes	10.3	13.8

*(Replace AAA and BBB with the actual airport codes, and fill in all the cells of the table.)*

### Method

I identified this route by running the following SELECT statement using Impala on the VM:

```
SELECT origin,  
       dest,  
       ROUND(AVG(distance), 1) AS avg_distance,  
       ROUND(COUNT(flight) /10) AS num_flights_per_year,  
       ROUND(SUM(seats)/10) AS capacity_per_year,  
       ROUND(AVG(arr_delay), 1) AS avg_arr_delay  
FROM flights f  
LEFT OUTER JOIN planes p ON f.tailnum = p.tailnum  
WHERE distance BETWEEN 300 AND 400  
GROUP BY origin,  
       dest  
HAVING num_flights_per_year >= 5000  
ORDER BY capacity_per_year DESC  
LIMIT 10;
```

(The result set of the query is the picture below)



fly

Tables

- airlines
- airports
- flights
- planes

(4) + -

Query History

Saved Queries

Results (10)

	origin	dest	avg_distance	num_flights_per_year	capacity_per_year	avg_arr_delay
1	SFO	LAX	337	14712	1996597	10.300000000000001
2	LAX	SFO	337	14540	1981059	13.800000000000001
3	PHX	LAX	370	8662	1219235	5.7999999999999998
4	LAX	PHX	370	8650	1210173	6
5	PHX	SAN	304	6200	1067278	4.9000000000000004
6	SAN	PHX	304	6216	1060204	3.7999999999999998
7	SLC	DEN	390.60000000000002	8012	920919	4.2000000000000002
8	DEN	SLC	390.60000000000002	7667	893437	6.2000000000000002
9	BOS	DCA	399	8484	867688	1.3999999999999999
10	DCA	BOS	399	8493	864009	4

javascript: void(0);