

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

**Name:** Abrar Hyder Mohammed

**Date:** 05/08/2022

*(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	PHX	LAX
Three-letter airport code for destination	LAX	PHX
Average flight distance in miles	370	370
Average number of flights per year	83970	83764
Average annual passenger capacity	1219235	1210173
Average arrival delay in minutes	6	6

*(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,
avg(distance) as avg_flight_dist,sum(seats/10) as avg_seat_capacity,

Count(flight/10) as num_flights_per_year, avg (arr_delay) as avg_arr_delay

from flights join planes using (talinum)
Where distance between 300 and 400
Group by origin,dest
Having num_flights_per_year >5000
Order by avg_seat_capacity desc nulls last
Limit 50;
```

*(Fill in the blank to indicate whether you used Hive or Impala and fill in the SQL query.)*

## Notes

*(This section is optional. You may use it to describe your process, add details or caveats, explain your interpretations, or describe any further analysis that you performed.)*

- *Since the two corresponding column name in join condition were equal, I used the “**using** keyword” to join the two tables instead the left inner join.*