Peer-Graded Assignment: Analyzing Big Data with SQL
Name: Süleyman Baver Özkeskin
Date: 09/03/2022

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	8397	8376
Average annual passenger capacity	1219235	1210173
Average arrival delay in minutes	6.08	6.03

Method

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

Notes

Below you can see a visual representation of the directions that had themost number of seat per year while having at least 5000 flights per year. It is pretty obvious that high speed passenger rail tunnel would be beneficial and efficient for Los Angeles International Airport (LAX) and Phoenix Sky Harbor International Airport (PHX).