

DSO 552: SQL Databases for Business Analysts

Homework 1

Due Date: Sunday September 8, 2019 (at the end of the day at 11:59pm)

Houston Flights

This database contains all flights departing from Houston airports IAH (George Bush Intercontinental) and HOU (Houston Hobby), and it has only one table called 'flights'. The data comes from the Research and Innovation Technology Administration at the Bureau of Transportation statistics: http://www.transtats.bts.gov/DatabaseInfo.asp?DB_ID=120&Link=0

The following explains the variables in the flights table:

Variable	Description
Year, Month, DayofMonth:	date of departure
DayOfWeek	day of week of departure (useful for removing weekend effects)
DepTime, ArrTime	departure and arrival times (in local time, recorded as hhmm)
UniqueCarrier	unique abbreviation for a carrier
FlightNum	flight number
TailNum	airplane tail number
ActualElapsedTime	elapsed time of flight, in minutes
AirTime	flight time, in minutes
ArrDelay, DepDelay	arrival and departure delays, in minutes
Origin, Dest	origin and destination airport codes
Distance	distance of flight, in miles
TaxiIn, TaxiOut	taxi in and out times in minutes
Cancelled	cancelled indicator: 1 = Yes, 0 = No
CancellationCode	reason for cancellation: A = carrier, B = weather, C = national air system, D = security
Diverted	diverted indicator: 1 = Yes, 0 = No

Loading the Data into PgAdmin

Here is a short video on how to read the data for this assignment in PgAdmin:

<https://usc-marshall-panopto-demo.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=9186efee-7bc8-4f7c-b6d0-aab7011ea>

1. (1 point) Which flight carriers fly out of Huston airports?

```
select distinct uniquecarrier
from flights;
```

2. (1 point) Select all delta and American Airlines flights.

```
select *
from flights
```

```
where uniquecarrier in ('DL', 'AA')
limit 5;
```

3. (1 point) Select all flights flying out of George Bush Intercontinental (IAH) airport with flying distance equal or larger than 3000 miles.

```
select *
from flights
where origin = 'IAH' AND distance >=3000
limit 5;
```

4. (1 point) Find the top 5 destinations for the flights flying out of George Bush Intercontinental (IAH) with the largest distance travelled. Report only the flight origin, destination, and distance.

```
select distinct origin, dest, distance
from flights
where origin = 'IAH'
order by distance desc
limit 5;
```

5. (1 point) Report all flights whose taxiing took longer than flying (taxiing = taxiIn + taxiOut). Report tailnum, taxiing, and airtime for those flights.

```
select tailnum, taxiin, taxiout, taxiin + taxiout as taxiing, airtime
from flights
where taxiin + taxiout > airtime
limit 5;
```

6. (1 point) Return all flights leaving to Dallas Fort Worth (DFW) airport before 8am sorted by descending air time.

```
select *
from flights
where dest = 'DFW' AND deptime < 800 AND AirTime is not null
order by airtime desc
limit 5;
```

7. (2 points) Which flights that were cancelled during the weekend? Report the carrier, day of the week, and cancellation status.

```
select uniquecarrier, dayofweek, cancelled
from flights
where dayofweek in (6,7) and cancelled = 1
limit 5;
```

8. (2 points) How many times American Airlines flights were delayed?

```
select count(*)
from flights
where uniquecarrier = 'AA' and arrdelay is not null
```