Worksheet 21 - writing SQL queries

Monday, April 21, 2025

DS 002R - Jo Hardin

Name: _____

Names of people you worked with: _____

What class are you most excited about for spring semester?

Task:

The airlines SQL database contains information on 18 million flights from the Bureau of Transportation Statistics (you have worked with a small subset of this data in the nycflights13 package). Information from the database can be obtained through SQL queries. For example, the flights table contains the following variables:

1. How many flights flew into Dallas-Fort Worth `DFW` on May 14, 2014?
2. Find all flights between `JFK` and `SFO` in 2015 How many were cancelled? What percentage of the total number of flights were cancelled? (Assume that `cancelled` is a binary variable with values `0` and `1`, so you can do arithmetic on it.)
3. Of all the destinations from Chicago O'Hare (`ORD`), which were the most common in 2013?
4. Which airport had the highest average arrival delay time in 2015?
5. How many flights came into or flew out of Bradley Airport (`BDL`) in 2013?
6. List the airline and flight number for all flights between `LAX` and `JFK` on September 26th, 2014.

```
DESCRIBE flights;
```

```
            Field        Type Null Key Default Extra
1            year smallint(4)  YES        <NA>
2           month smallint(2)  YES        <NA>
3             day smallint(2)  YES        <NA>
4        dep_time smallint(4)  YES        <NA>
5  sched_dep_time smallint(4)  YES        <NA>
6       dep_delay smallint(4)  YES        <NA>
7        arr_time smallint(4)  YES        <NA>
8  sched_arr_time smallint(4)  YES        <NA>
9       arr_delay smallint(4)  YES        <NA>
10        carrier  varchar(2)   NO
11        tailnum  varchar(6)  YES MUL    <NA>
12         flight smallint(4)  YES        <NA>
13         origin  varchar(3)   NO MUL
14           dest  varchar(3)   NO MUL
15       air_time smallint(4)  YES        <NA>
16       distance smallint(4)  YES        <NA>
17      cancelled  tinyint(1)  YES        <NA>
18       diverted  tinyint(1)  YES        <NA>
19           hour smallint(2)  YES        <NA>
20         minute smallint(2)  YES        <NA>
21      time_hour    datetime  YES        <NA>
```

Solution:

1. How many flights flew into Dallas-Fort Worth `DFW` on May 14, 2014?

```
SELECT sum(1) AS N
FROM flights
WHERE year = 2014 AND Month = 5 AND day = 14
  AND dest = 'DFW';
```

Table 1: 1 records

N
786

2. Find all flights between `JFK` and `SFO` in 2015 How many were cancelled? What percentage of the total number of flights were cancelled?

```
SELECT
  SUM(1) AS numFlights,
  SUM(cancelled) AS numCancelled,
  AVG(cancelled) AS pctCancelled
FROM flights
WHERE year = 2015
  AND dest IN ('JFK', 'SFO') AND origin IN ('JFK', 'SFO')
LIMIT 0,6;
```

Table 2: 1 records

numFlights	numCancelled	pctCancelled
18559	230	0.0124

3. Of all the destinations from Chicago O'Hare (`ORD`), which were the most common in 2013?

```
SELECT
  dest,
  SUM(1) AS numFlights
FROM flights
WHERE year = 2013
  AND origin = 'ORD'
```

```
GROUP BY dest
ORDER BY numFlights desc
LIMIT 0,6;
```

Table 3: 6 records

dest	numFlights
LGA	8876
SFO	7950
LAX	7799
DFW	7613
ATL	7584
BOS	6969

4. Which airport had the highest average arrival delay time in 2015?

```
SELECT
  dest,
  SUM(1) AS numFlights,
  AVG(arr_delay) AS avg_arr_delay
FROM flights
WHERE year = 2015
GROUP BY dest
ORDER BY avg_arr_delay desc
LIMIT 0,6;
```

Table 4: 6 records

dest	numFlights	avg_arr_delay
STC	82	21.6220
ILG	98	21.3163
OTH	311	17.0868
TTN	3113	15.7600
PBG	295	15.3559
GUM	365	14.3671

5. How many flights came into or flew out of Bradley Airport (`BDL`) in 2013?

```sql
SELECT COUNT(*) AS numFlights
FROM flights
WHERE Year = 2013
  AND (Dest = 'BDL' OR Origin = 'BDL');
```

Table 5: 1 records

numFlights
44516

6. List the airline and flight number for all flights between `LAX` and `JFK` on September 26th, 2014.

```sql
SELECT carrier, flight
FROM flights
WHERE Year = 2014 AND Month = 9 AND day = 26
  AND Origin IN ('LAX', 'JFK')
  AND Dest IN ('LAX', 'JFK')
LIMIT 0,6;
```

Table 6: 6 records

carrier	flight
VX	420
AA	292
UA	212
DL	476
AA	118
B6	24