1. Use the STR_TO_DATE() function to further refine your solution to problem 9 on the first assignment, such that the date column is recognized as a date by MySQL. The resulting column should be formatted in ISO 8601. (2 points)

select str_to_date(concat(year,'-',month,'-',day), "%Y-%m-%d") as the_date, origin, dest, flight, carrier from flights limit 0,10;

2. Display only those flights that left John F. Kennedy airport in New York City and landed in Los Angeles International airport on April 8, 2014. (1 point)

select str_to_date(concat(year,'-',month,'-',day), "%Y-%m-%d") as the_date, flights.* from flights where flights.origin='JFK' and flights.dest='LAX' and str_to_date(concat(year,'-',month,'-',day), "%Y-%m-%d")='2014-04-08';

3. Count the number of flights scheduled to land at San Francisco International airport during each month of 2013. (1 point)

select year, month, count(*) as total from flights where dest='SFO' and year=2013 group by year, month;

4. Repeat the previous query, but now add a column that counts how many of those scheduled flights were cancelled. (2 points)

select year, month, count(*) as total, sum(if(cancelled=1, 1, 0)) as cancelledtotal from flights

where dest='SFO' and year=2013 group by year, month;

5. Repeat the previous query, but now add a column that computes the percentage of cancelled flights in each month. Arrange the results in descending order by cancellation percentage. (3 points)

select year, month, count(*) as total, sum(if(cancelled=1, 1, 0)) as cancelledtotal,

concat(sum(if(cancelled=1, 1, 0))/count(*) *100, '%') as cancelledpercent from flights

where dest='SFO' and year=2013 group by year, month order by sum(if(cancelled=1, 1, 0))/count(*) desc;

6. Find the tail number of the plane that flew the most flights in 2015. Include only planes with at least 3500 segments in your results. (3 points)

select year, tailnum, count(flight) as totalflight from flights where year=2015 group by tailnum having count(flight)>=3500 order by totalflight desc;

7. US airlines generally have regional hubs that process many flights. List the top 10 airports used by American Airlines (for departing flights only) in 2013. Can you identify the hubs? (3 points)

select year, carrier, origin, count(*) as totalflights from flights where carrier='AA' and year=2013 group by origin order by totalflights desc limit 10;

8. Southwest Airlines, on the other hand, does not employ the hub-and-spoke model for airlines traffic. Compute the top 10 airports used (for departing flights only) by Southwest Airlines in 2013. Do you notice a difference in the distribution of the number of flights across airports as compared to American? (3 points)

select year, carrier, origin, count(*) as totalflights from flights where carrier='WN' and year=2013 group by origin order by totalflights desc limit 10;

9. Find the busiest airports (most departures) in 2015 across all carriers. (3 points)

select year, origin, count(*) as totalflights from flights where year=2015 group by origin order by totalflights desc limit 10;

10. Among all non-cancelled flights in 2014, list the destination airports in ascending order of their average arrival delay time. Include only airports that had at least 50,000 arrivals. (3 points)

select year, dest, count(*) as totalarrivals, sum(arr_delay)/count(*) as avgarrdelay from flights where cancelled=0 and year=2014 group by dest having count(*)>=50000 order by avgarrdelay;

11. List all of the airports that have the word Chicago in their name. (1 point)

select * from airports where name like '%Chicago%';