Part 1 – SQL

Given the below subset of a travel app's schema, write executable SQL queries to answer the two questions below. Please answer in a single query and assume read-only access to the database (i.e. do not use CREATE TABLE).

Assume a PostgreSQL database, server timezone is UTC.

Table Name: trips

Column Name:	Datatype:
id	integer
client_id	integer (Foreign keyed to users.usersid)
driver_id	integer (Foreign keyed to users.usersid)
city_id	integer
client_rating	integer
driver_rating	integer
status	Enum('completed', 'cancelled_by_driver', 'cancelled_by_client')
actual_eta	integer
request_at	timestamp with timezone

Table Name: users

Column Name:	Datatype:
usersid	integer
email	character varying
signup_city_id	integer
banned	Boolean
role	Enum('client', 'driver', 'partner')
created_at	timestamp with timezone

Question 1

Q: Between Oct 1, 2013 at 10am PDT and Oct 22, 2013 at 5pm PDT, what percentage of requests made by unbanned clients each day were canceled in each city?

A:

```
SELECT  city_id,
        date_id,
        /* Count total cancelled trips*/
        Sum(
                CASE WHEN status = 'completed' THEN 0
                     ELSE 1
                END) / Count(*) :: float AS Percentage_Cancelled
/*Join the trips table and the users table*/
FROM
    /*Select all trips within the time range*/
    (
        SELECT client_id,
        city_id,
        status,
        date_trunc('day', request_at AT TIME ZONE 'PDT') AS date_id
        FROM trips
        WHERE request_at >= TIMESTAMP WITH TIME ZONE '2013-10-01 10:00:00 PDT'
        AND request_at <= TIMESTAMP WITH TIME ZONE '2013-10-22 17:00:00 PDT'
    )   total_trips
INNER JOIN
    /* Select unbanned clients*/
    (
        SELECT usersid
        FROM users
        WHERE banned = FALSE
        AND role = 'client'
    )   unbanned_clients
ON      total_trips.client_id = unbanned_clients.usersid
GROUP BY
        city_id,
        date_id;
```

Question 2
Q: For city ids 1, 6, and 12, list the top three drivers by number of completed trips for each week between June 3, 2013 and June 24, 2013.

A:
NOTE: Assumed that weeks are supposed to be separated based on real calendar information and each week starts from Sunday and ends on Saturday. Therefore, there were 4 weeks within the given time range with the first week been from June 3, Monday to June 8, Saturday and the last week been from June 23, Sunday to June 24, Monday (included).

Assumed by saying '*For city ids 1, 6, 12'*, the question was meant to say *'List the top three drivers by number of completed trips for each week between June 3, 2013 and June 24, 2013 separately for city ids 1, 6, 12'* for which ignores the fact that a driver might have completed trips in multiple cities and treat the same driver in different cities as different individuals.

```
SELECT city_id,
       driver_id,
       /* Assign week_id*/
       (
            CASE WHEN request_at >= '2013-06-03' AND request_at <= '2013-06-08'
            THEN 'week 1'
            WHEN request_at >= '2013-06-09' AND request_at <= '2013-06-15'
            THEN 'week 2'
            WHEN request_at >= '2013-06-16' AND request_at <= '2013-06-22'
            THEN 'week 3'
            ELSE 'week 4'
            END
       )week_id,
       Count(*) AS trips_count
FROM trips
/* Condition time range between June 3, 2013 and June 24, 2013*/
WHERE request_at >= '2013-06-03'
AND request_at <= '2013-06-24'
/* Condition city ids 1, 6, 12*/
AND city_id IN (1, 6, 12)
/* Condition 'completed trips'*/
AND status = 'completed'
GROUP BY city_id,
         week_id,
         driver_id
ORDER BY trips_count desc limit 3
```

Self-Assessment
Q: On a scale of 1-5 with 5 being for a perfect answer, where do you think your answer stands?

A: 4/5. Further SQL Optimization can be made to improve running time based on needs.