**SQL TEST:**

**Background:** schema, datatypes, and schema relationships:

**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 |
| request\_device | Enum(‘android’, ‘iphone’, ‘sms’, ‘mobile\_web’) |
| status | Enum(‘completed’, ‘cancelled\_by\_driver’, ‘cancelled\_by\_client’) |
| predicted\_eta | integer |
| actual\_eta | integer |
| request\_at | timestamp with timezone |

Table Name: **users**

|  |  |
| --- | --- |
| Column Name: | Datatype: |
| usersid | integer |
| email | character varying |
| firstname | character varying |
| lastname | character varying |
| banned | Boolean |
| role | Enum(‘client’, ‘driver’, ‘partner’) |
| creationtime | timestamp with time zone |

**Problem Statement:** It’s common at Uber to want to know various business metrics about recent trips. Given the above subset of Uber’s schema, write executable SQL queries to answer the following questions:

1. For request times between 12/1/2013 10:00:00 PST & 12/8/2013 17:00:00 PST, how many completed trips (Hint: look at the trips.status column) were requested on iphones in City #5? on android phones?
2. In City #8, how many unique, currently unbanned clients requested a trip in October 2013 that was eventually completed? Of these, how many trips did each client take?
3. In City #8, how many unique, currently unbanned clients requested a trip between 9/10/2013 and 9/20/2013, with drivers who started between 9/1/2013 and 9/10/2013 and are currently banned, that was eventually completed?

**Extra Credit:** Add to your statement in 2) to exclude Uber admins. Uber admins have an email address from @uber.com (example: ‘[jsmith@uber.com](mailto:jsmith@uber.com)’).

**Data Analysis Challenge:**

We’ve attached a JSON dataset of client logins from an Uber city on the eastern seaboard of the United Statues. Using this, please do the following:

1. Using your analysis tool of choice, please generate a graph showing an hourly breakdown of client login behavior.
2. Add a best fit line to this graph, and include any relevant metrics/statistics to quantify the quality of fit.
3. In a short (1-5 page) writeup, discuss any significant trends or deviations you observe about the dataset.

You may use your analysis tool of choice, but please include a short description of your method and source code, if appropriate.

**Extra Credit**: Repeat this analysis by graphing client logins by week and/or by hour of day. What do you notice about client behavior?