# Part 1

Nota Bene: I’ve made a few assumptions, such as what time best represents a trip, and that users do not have multiple completed sign ups. My decisions are heuristic, and normally would be confirmed with existing staff or the product owner before releasing analysis.

Additionally, I’ve put little emphasis in manual query optimization. For example, for question 2 could subset sub-tables for the for the first week of 2016. I assume that the query planner used is smart enough to make these optimizations.

**Question 1**

**Q:** For each of the cities 'Qarth' and 'Meereen', calculate 90th percentile difference between Actual and Predicted ETA for all completed trips within the last 30 days.

**A:**

**SELECT** PERCENTILE\_CONT(.9)  
 WITHIN **GROUP** (**ORDER BY** trips.actual\_eta-predicted\_eta)  
 **AS** 90th\_percentile  
 **FROM** trips  
 **LEFT OUTER JOIN** cities  
 **WHERE** trips.city\_id == cities.city\_id  
 **WHERE** cities.city\_name **IN** ('Qarth', 'Meereen')  
 **AND** trips.status == 'completed'  
 **AND** trips.request\_at > (**CURRENT\_TIMESTAMP**- **INTERVAL** '10 days');

**Question 2**

**Q:** A signup is defined as an event labeled ‘sign\_up\_success’ within the events table. For each city (‘Qarth’ and ‘Meereen’) and each day of the week, determine the percentage of signups in the first week of 2016 that resulted in completed a trip within 168 hours of the sign up date.

**A:**

**SELECT** signups\_enhanced.day\_of\_week, *AVG*(rode\_in\_first\_week::**int**)  
 **FROM** -- Create sub-table with one row for every rider who signed up, with rode\_in\_first\_week metric  
 ( **SELECT** events.\*  
 **EXTRACT**( **DOW FROM** \_ts) **AS** day\_of\_week  
 -- Actually compute rode\_in\_first\_week metric  
 -- Check if user has a ride  
 (*MIN*(trips.request\_at) **IS NOT NULL** -- First ride within 168 hours  
 **AND** *MIN*(trips.request\_at) <= *MIN*(events.\_ts) + **INTERVAL** '168 hours'  
 -- No rides before sign up  
 **AND** *MIN*(trips.request\_at) >= *MIN*(events.\_ts))  
 **AS** rode\_in\_first\_week  
 **FROM** trips  
 **LEFT OUTER JOIN** -- Create sub-table with every rider's first completed trip  
 (**SELECT DISTINCT ON** (trips.client\_id) trips.client\_id, request\_at  
 **FROM** trips  
 **WHERE** status == 'completed'  
 ) **AS** first\_completed\_trips  
  
 **WHERE** events.rider\_id == first\_completed\_trips.client\_id  
 **AND** event\_name == 'sign\_up\_success'  
 ) **AS** signups\_enhanced  
  
 **GROUP BY** signups\_enhanced.day\_of\_week  
 **WHERE EXTRACT**(**WEEK FROM** signup\_ts) == 1  
 **AND EXTRACT**(**YEAR FROM** signup\_ts) == 2016;  
 **AND** city\_name **IN** ('Qarth', 'Meereen');