# 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:**

This query is somewhat contrived. If this is a common data access pattern, I would normally work with the product owner to understand if the sub-tables I generate (*first completed ride timestamp* and *rode in first week*) are valuable enough to be stored on their own, or if there is a more optimal way to store this data in a more accessible way.

**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** trips.status == 'completed'  
 **ORDER BY** trips.request\_at **ASC** ) **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');

# Part 2

**Question 1**

**Q:**

Propose and define the primary success metric of the redesigned app. What are 2­3 additional tracking metrics that will be important to monitor in addition to the success metric defined above?

**A:**

Ideally, during planning of the new release I would work with the team behind the new release to identify their goals and metrics that capture them.

I would propose the following metrics:

* **New feature time:** The amount of time spent in the four new drive app sections (Home, Earnings, Ratings, Account )
* **Driver Productivity:** The difference in total fares seen by drivers who use the new app and to drivers who use the old app
* **Driver help requests:** The difference in driver contacts (e.g. email, phone) to Uber between drivers who use the new app and drivers who use the old app

Additionally, before the trial begins I would work with the product team to choose acceptable thresholds for each metric for a new release. For example, we might use the thresholds below, with a pre-defined statistical siginificance:

|  |  |
| --- | --- |
| Metric | Threshold |
| New feature time | 20 minutes a week or more |
| Driver productivity | No change or increase |
| Driver help requests | No change or decrease |

**Question 2**

**Q:**

Outline a testing plan to evaluate if redesigned app performs better (according to the metrics you outlined). How would you balance the need to deliver quick results, with statistical rigor, and while still monitoring for risks?

**A:**

*Existing protocols*

First, I would reach out to the team that is commonly tasked with A/B testing the rider app, and seek their recommendations or general testing framework. This would provide consistency in mobile A/B testing, and reduce redundant work in developing testing frameworks.

*Assumptions*

Additionally, I would confirm a few assumptions:

* Driver app versions before the new release are not substantially different
* Driver apps are not substantially different based on operating system (for both the current release and the new release)
* The metrics above meet product owner needs

*Test design*

In the absence of a testing plan from the rider app team, I would proceed with the following trial:

Segmentation: 3 distinct geographic locations, with 25% of drivers in each location forcibly upgraded to new version, and the remaining 75% forcibly frozen in their current version

Length: 1 month

Safety checks: Hourly checks for drop in driver productivity (in case of app bugs or crashes), and monitoring driver help requests (to ease burn in period, watch for bugs or crashes)

*Summary*

I feel that this trial design would be large and diverse enough to capture meaningful signal, without unduly exposing a large population of drivers to an unproven re-design. Additionally, network effects (e.g. drivers with the old app seeing drivers with the new app) should be minimal, and could be controlled by branding the new release as a ‘pre-release’ version. Finally, a one month test period should be enough to gather statistically significant results, and quickly iterate on the testing framework.

**Question 3**

**Q:**

Explain how you would translate the results from the testing plan into a decision on whether to launch the new design or roll it back.

**A:**

I would evaluate each of the metrics designed at the onset of the trial, relative to the thresholds designated at the onset of the trial.

I would then identify if the thresholds and statistical significance levels were appropriate, and adjust the thresholds and statistical significance levels as appropriate.