Part 2: Experiment and Metrics Design

1. What would you choose as the key measure of success of this experiment in encouraging driver partners to serve both cities, and why would you choose this metric?

The best metric for success would be the percentage of drivers who cross the bridge from the city with less demand to the city with greater demand when there is a demand imbalance in a given period of time. To perform the experiment properly, we would want to know with high confidence what percentage of drivers cross the bridge when Ultimate doesn't pay the toll so we can see how much this increases when Ultimate pays the toll. We would want this increase to be as large as possible.

A secondary, related metric we could use is the time it takes to even out demand imbalances. We would expect it to be smaller when we pay the tolls than when we don't. A number of assumptions were made in the above answer. If we are in a more naïve world where Ultimate doesn't have revenue maximization as its primary goal and its primary goal in this case is really to have drivers crossing over often, then the best measure of success would be the most obvious one: the number of times drivers cross the bridge per whatever time frame we want. This is equivalent to the expenditures Ultimate must make for paying the toll. In this more naïve world, Ultimate wants to maximize trips made and its expenditures on tolls.

2. Describe a practical experiment you would design to compare the effectiveness of the proposed change in relation to the key measure of success. Please provide details on:

a. How you will implement the experiment

As mentioned in the first answer, before implementing the proposed change (paying drivers' toll fees) we would want to get a really strong estimate of what percentage of drivers already cross the bridge and pay the toll in order to take advantage of the higher demand in the other city.
We would then want to see what that number changes to when we pay the toll fees.
In terms of practical implementation, this should not be difficult to implement assuming Ultimate has an app that tracks drivers locations as well as the current average demand in each city (as given by price/unit distance).

b. What statistical test(s) you will conduct to verify the significance of the observation

We would use a hypothesis test comparing population proportions where the proportion in question is the percentage of drivers who cross from the city with less demand to the city with greater demand in a given frame of time. Our null hypothesis would be that the proportion is the same irrespective of our proposed change.

c. How you would interpret the results and provide recommendations to the city operations team along with any caveats.

If we saw that the change in proportion (or percentage) of drivers who cross the bridge from the city with greater demand to the city with less demand in a given period was large, than we would know that paying the drivers' toll fees was an effective intervention, at least in terms of meeting demand quicker. If we saw that this percentage change was small, then we would know that the proposed change was not effective, and we would need to provide additional incentives to drivers to cross the bridge.

My recommendations to the operations team would depend a lot on some of the considerations already mentioned above. That is, does demand remain strong enough in the city to which drivers are going and do drivers perform enough trips in the 'new' city to justify the cost of having Ultimate pay the toll? Ideally there should be some analysis performed on this question before the change is implemented to know whether the change should be made in the first place. But some questions can only be answered with real data, so from my perspective, as a consultant tasked with maximizing Ultimate's revenue, I would want to know if the net effect to revenue from paying the tolls before recommending to city managers whether to keep the change or not.